THE UNIVERSITY OF QUEENSLAND

AUSTRALIA

Protein expression and molecular profiling to predict lymph node status and prognosis in breast cancer.

Glenn Duval Francis

MBBS FRCPA MBA FFSc (RCPA)

A thesis submitted for the degree of Doctor of Philosophy at

The University of Queensland in 2015

School of Dentistry

Abstract

Introduction: The hypothesis of this study is that prognostic and predictive markers defined by immunohistochemistry (IHC) and other histopathologic characteristics of a primary breast carcinoma tumour, together with gene expression profiling of the tumour, DNA methylation profiling and microRNA (miRNA) will identify lymph node status in breast cancer patients and allow selection of patients for sentinel lymph node biopsy (SLNB) or axillary lymph node dissection (ALND) in addition to providing prognostic information.

Methods: A web-based database of breast cancer patients was constructed using Distiller, SlidePath (Leica Microsystems, Germany) software. A MindMap was used to determine subgroupings and structure of the database. Three patient cohorts were identified from separate sources and patient records were identified using Systematized Nomenclature of Medicine (SNOMED) coding as a search parameter. Data was extracted into Microsoft Excel and uploaded into Distiller.

3200 patient reports were retrieved from a private pathology database, 1834 from a public hospital and 5200 from a separate public pathology hospital. Haematoxylin and eosin (H&E) slides from 645 patients were reviewed from the initial dataset to confirm the imported data and reliability of pathology information. This included tumour type, tumour grade and size, presence or absence of lymphatic and vascular permeation, margin type, lymphocytic infiltration, lymph node status and number of involved lymph nodes. Survival data was also obtained from the Queensland Cancer Registry under HREC approval to evaluate applications of the data for prognostic information. A breast cancer cluster was incorporated into the project and evaluated to ascertain if there were commonalities between the tumours which would offer potential insight into causes of breast cancer or prognostic information.

Tissue microarrays (TMAs) were constructed to enable evaluation of immunohistochemistry markers and TMAs were evaluated for correlation with clinical parameters. Epigenetic biomarkers were also evaluated and correlated with different parameters in triple negative breast carcinoma (TNBC).

Results: Biomarkers evaluated on TMAs successfully correlated with clinical outcome in breast carcinoma patients and an artificial neural network (ANN) model was constructed to predict modular grade in breast cancer patients. This ANN model enabled the subdivision of Grade 2 breast carcinomas into Grade 1 and Grade 3 carcinomas and correlated with patient survival. Epigenetics models were also able to stratify TNBC patients into prognostic groups.

However, an ANN was not able to be successfully developed to predict lymph node status and similarly epigenetics was also not able to predict lymph node status. An analysis of a breast cancer cluster also showed that the cluster tumours were similar to sporadic breast cancers and a distinctive pattern was not identified.

The project has enabled the identification of markers to subclassify breast carcinomas into a molecular grade and to develop a number of immunohistochemistry and epigenetic biomarkers that correlate with prognosis in breast cancer.

However, the data from this project suggests that it is not possible to predict lymph node status form the primary tumour characteristics with sufficient clinical accuracy to forgo sentinel lymph node biopsy. It also suggests that lymph node metastases and distant metastases are separate independent events. Lymph node metastases are associated with a complex phenotype that is not defined a single genetic mutation and the elucidation of the events resulting in this biological cascade are still to be defined.

Declaration by author

This thesis is composed of my original work, and contains no material previously published or written by another person except where due reference has been made in the text. I have clearly stated the contribution by others to jointly-authored works that I have included in my thesis.

I have clearly stated the contribution of others to my thesis as a whole, including statistical assistance, survey design, data analysis, significant technical procedures, professional editorial advice, and any other original research work used or reported in my thesis. The content of my thesis is the result of work I have carried out since the commencement of my research higher degree candidature and does not include a substantial part of work that has been submitted to qualify for the award of any other degree or diploma in any university or other tertiary institution. I have clearly stated which parts of my thesis, if any, have been submitted to qualify for another award.

I acknowledge that an electronic copy of my thesis must be lodged with the University Library and, subject to the policy and procedures of The University of Queensland, the thesis be made available for research and study in accordance with the Copyright Act 1968 unless a period of embargo has been approved by the Dean of the Graduate School.

I acknowledge that copyright of all material contained in my thesis resides with the copyright holder(s) of that material. Where appropriate I have obtained copyright permission from the copyright holder to reproduce material in this thesis.

Publications during candidature

Peer reviewed papers.

1. FRANCIS, G. D., JONES, M. A., BEADLE, G. F. & STEIN, S. R. 2009. Bright-field in situ hybridization for HER2 gene amplification in breast cancer using tissue microarrays: correlation between chromogenic (CISH) and automated silver-enhanced (SISH) methods with patient outcome. Diagn Mol Pathol, 18, 88-95. Wolters Kluwer Health Lippincott Williams & Wilkins©.
2. WEE, E. J., PETERS, K., NAIR, S. S., HULF, T., STEIN, S., WAGNER, S., BAILEY, P., LEE, S. Y., QU, W. J., BREWSTER, B., FRENCH, J. D., DOBROVIC, A., FRANCIS, G. D., CLARK, S. J. & BROWN, M. A. 2012. Mapping the regulatory sequences controlling 93 breast cancer-associated miRNA genes leads to the identification of two functional promoters of the Hsa-mir-200b cluster, methylation of which is associated with metastasis or hormone receptor status in advanced breast cancer. Oncogene, 31, 4182-95.
3. WADDELL, N., STEIN, S. R., WAGNER, S. A., BENNETT, I., DJOUGARIAN, A., MELANA, S., JAFFER, S., HOLLAND, J. F., POGO, B. G., GONDA, T. J., BROWN, M. A., LEO, P., SAUNDERS, N. A., MCMILLAN, N. A., COCCIARDI, S., VARGAS, A. C., LAKHANI, S. R., CHENEVIX-TRENCH, G., NEWMAN, B. & FRANCIS, G. D. 2012. Morphological and molecular analysis of a breast cancer cluster at the ABC Studio in Toowong. Pathology, 44, 469-72. Wolters Kluwer Health Lippincott Williams & Wilkins©.
4. PETERS, K. M., EDWARDS, S. L., NAIR, S. S., FRENCH, J. D., BAILEY, P. J., SALKIELD, K., STEIN, S., WAGNER, S., FRANCIS, G. D., CLARK, S. J. & BROWN, M. A. 2012. Androgen receptor expression predicts breast cancer survival: the role of genetic and epigenetic events. BMC Cancer, 12, 132.
5. FRANCIS, G. D., STEIN, S. R. Prediction of histologic grade in breast cancer using an artificial neural network. Neural Networks (IJCNN), The 2012 International Joint Conference on, 10-15 June 2012 2012. 1-5. © 2012 IEEE. Reprinted, with permission. In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of the Univeristy of Queenslnad's products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to

http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.

6. PLANT, H. C., KASHYAP, A. S., MANTON, K. J., HOLLIER, B. G., HURST, C. P., STEIN, S. R., FRANCIS, G. D., BEADLE, G. F., UPTON, Z. & LEAVESLEY, D. I. 2014. Differential subcellular and extracellular localisations of proteins required for insulin-like growth factor- and extracellular matrix-induced signalling events in breast cancer progression. BMC Cancer, 14, 627.
7. STIRZAKER, C., ZOTENKO, E., SONG, J. Z., QU, W., NAIR, S. S., LOCKE, W. J., STONE, A., ARMSTONG, N. J., ROBINSON, M. D., DOBROVIC, A., AVERY-KIEJDA, K. A., PETERS, K. M., FRENCH, J. D., STEIN, S., KORBIE, D. J., TRAU, M., FORBES, J. F., SCOTT, R. J., BROWN, M. A., FRANCIS, G. D. & CLARK, S. J. 2015. Methylome sequencing in triple-negative breast cancer reveals distinct methylation clusters with prognostic value. Nat Commun, 6, 5899.

Publications included in this thesis

FRANCIS, G. D., JONES, M. A., BEADLE, G. F. & STEIN, S. R. 2009. Bright-field in situ hybridization for HER2 gene amplification in breast cancer using tissue microarrays: correlation between chromogenic (CISH) and automated silver-enhanced (SISH) methods with patient outcome. Diagn Mol Pathol, 18, 88-95. – incorporated as Chapter 3.

Contributor	Statement of contribution
Author FRANCIS G.D. (Candidate)	Concept of project 100% Designed experiments (90%) Methods (70%) Reviewed Slides (100%) Statistical analysis (80%) Wrote and edited the paper (60%)
Author JONES. M.A.	Methods (10%) Statistical analysis (20%) Wrote and edited paper (5%)
Author STEIN S. R.	Designed experiments (10%) Construction of TMAs (100%)

	Methods (20%) Wrote and edited paper (30%)
Author BEADLE G. F.	Edited paper 5%

WEE, E. J., PETERS, K., NAIR, S. S., HULF, T., STEIN, S., WAGNER, S., BAILEY, P., LEE, S. Y., QU, W. J., BREWSTER, B., FRENCH, J. D., DOBROVIC, A., FRANCIS, G. D., CLARK, S. J. & BROWN, M. A. 2012. Mapping the regulatory sequences controlling 93 breast cancer-associated miRNA genes leads to the identification of two functional promoters of the Hsa-mir-200b cluster, methylation of which is associated with metastasis or hormone receptor status in advanced breast cancer. Oncogene, 31, 4182-95. – incorporated as Chapter 4.

Contributor	Statement of contribution
Author FRANCIS G. D. (Candidate)	Designed experiments (20%) Methods including slide review and scoring of TMAs (20%) Wrote and edited the paper (20%)
Author WEE E.J.	Designed experiments (40%) Methods and analysis (40%) Wrote and edited paper (30%)
Author PETERS, K., NAIR, S. S., HULF, T., BAILEY, P., LEE, S. Y., QU, W. J., BREWSTER, B., FRENCH, J. D., DOBROVIC, A., CLARK, S. J. & BROWN, M. A.	Designed experiments (40%) Methods and Statistical analysis of data (35%) Wrote and edited paper (50%)
Author STEIN S. & WAGNER S.	Methods including construction of TMAs (5%)

WADDELL, N., STEIN, S. R., WAGNER, S. A., BENNETT, I., DJOUGARIAN, A., MELANA, S., JAFFER, S., HOLLAND, J. F., POGO, B. G., GONDA, T. J., BROWN, M. A., LEO, P., SAUNDERS, N. A., MCMILLAN, N. A., COCCIARDI, S., VARGAS, A. C., LAKHANI, S. R., CHENEVIX-TRENCH, G., NEWMAN, B. & FRANCIS, G. D. 2012. Morphological and molecular analysis of a breast cancer cluster at the ABC Studio in Toowong. Pathology, 44, 469-72. – incorporated as Chapter 5.

Contributor	Statement of contribution
Author FRANCIS G.D. (Candidate)	Concept of project (90%) Designed experiments (40%) Methods and analysis (30%) Wrote the paper (30%)
Author WADDELL N.	Designed experiments (40%) Methods and analysis (40%) Wrote and edited paper (50%)
Author STEIN, S. R.& WAGNER, S.	Methods including TMAs (5%)
Author BENNETT, I., DJOUGARIAN, A., MELANA, S., JAFFER, S., HOLLAND, J. F., POGO, B. G., GONDA, T. J., BROWN, M. A., LEO, P., SAUNDERS, N. A., MCMILLAN, N. A., COCCIARDI, S., VARGAS, A. C., CHENEVIX-TRENCH, G., NEWMAN, B	Concept of project (10%) Designed experiments (10%) Methods (10%) Wrote and edited paper (15%)
Author LAKHANI, S.	Designed experiments (10%) Methods (15%) Wrote and edited paper (5%)

PETERS, K. M., EDWARDS, S. L., NAIR, S. S., FRENCH, J. D., BAILEY, P. J., SALKIELD, K., STEIN, S., WAGNER, S., FRANCIS, G. D., CLARK, S. J. & BROWN, M. A. 2012. Androgen receptor expression predicts breast cancer survival: the role of genetic and epigenetic events. BMC Cancer, 12, 132. – incorporated as Chapter 6.

Contributor	Statement of contribution
Author FRANCIS G. D. (Candidate)	Designed experiments (20%) Methods including slide review and scoring of TMAs (20%) Wrote and edited the paper (20%)
Author PETERS K.M.	Designed experiments (40%) Methods and analysis (40%) Wrote and edited paper (30%)

Author EDWARDS, S. L., NAIR, S. S., FRENCH, J. D., BAILEY, P. J., SALKIELD, K., CLARK, S. J. & BROWN, M. A.	Designed experiments (40%) Methods and Statistical analysis of data (35%) Wrote and edited paper (50%)
Author STEIN S. & WAGNER S.	Methods including construction of TMAs (5%)

FRANCIS, G. D., STEIN, S. R. Prediction of histologic grade in breast cancer using an artificial neural network. Neural Networks (IJCNN), The 2012 International Joint Conference on, 10-15 June 2012 2012. 1-5. – incorporated as Chapter 7.

Contributor	Statement of contribution
Author FRANCIS G.D. (Candidate)	Concept of project (90%) Designed experiments (90%) Methods including scoring of TMAs and construction of ANN (90%) Statistical analysis of data (100%) Wrote the paper (80%)
Author STEIN S. R.	Concept of project (10%) Designed experiments (10%) Methodology (10%) Wrote and edited paper (20%)

PLANT, H. C., KASHYAP, A. S., MANTON, K. J., HOLLIER, B. G., HURST, C. P., STEIN, S. R., FRANCIS, G. D., BEADLE, G. F., UPTON, Z. & LEAVESLEY, D. I. 2014. Differential subcellular and extracellular localisations of proteins required for insulin-like growth factor- and extracellular matrix-induced signalling events in breast cancer progression. BMC Cancer, 14, 627. – incorporated as Chapter 8.

Contributor	Statement of contribution
Author FRANCIS G.D. (Candidate)	Designed experiments (20%) Methods including slide review and scoring of TMAs (50%) Wrote and edited the paper (5%)
Author PLANT, H. C.	Designed experiments (40%)

	Methods including analysis (30%) Wrote and edited paper (50%)
Author STEIN S.	Methods including construction of TMAs (5%)
Author KASHYAP, A. S., MANTON, K. J., HOLLIER, B. G., HURST, C. P., BEADLE, G. F., UPTON, Z. & LEAVESLEY, D. I.	Designed experiments (40%) Methods and Statistical analysis of data (15%) Wrote and edited paper (45%)

STIRZAKER, C., ZOTENKO, E., SONG, J. Z., QU, W., NAIR, S. S., LOCKE, W. J., STONE, A., ARMSTONG, N. J., ROBINSON, M. D., DOBROVIC, A., AVERY-KIEJDA, K. A., PETERS, K. M., FRENCH, J. D., STEIN, S., KORBIE, D. J., TRAU, M., FORBES, J. F., SCOTT, R. J., BROWN, M. A., FRANCIS, G. D. & CLARK, S. J. 2015. Methylome sequencing in triple-negative breast cancer reveals distinct methylation clusters with prognostic value. Nat Commun, 6, 5899. – incorporated as Chapter 9.

Contributor	Statement of contribution
Author FRANCIS G.D. (Candidate)	Concept of project (25%) Designed experiments (10%) Methods including scoring of TMAs (20%) Analysis and Interpretation of data (10%) Wrote and edited the paper (10%)
Author DOBROVIC, A., TRAU, M., FORBES, J. F., SCOTT, R. J., BROWN, M. A., & CLARK, S. J.	Concept of project (75%)
Author QU, W., AVERY-KIEJDA, K. A., PETERS, K. M., FRENCH, J. D., STEIN, S.	Preparation of DNA (100%)
Author ZOTENKO, E., LOCKE, W. J., STONE, A., ARMSTONG, N. J., ROBINSON, M. D., NAIR, S. S.	Analysis and interpretation of the data (90%)
Author STIRZAKER, C., ZOTENKO, E., SONG, J. Z., LOCKE, W. J., STONE, A., KORBIE, D. J., CLARK, S. J.	Designed experiments (90%) Methods (80%) Writing, figures, review of the manuscript (90%)

Contributions by others to the thesis

Contributions of others to the thesis are detailed in the Publications list included in the thesis for Chapters 3 – 9.
Contributions for methodology include Stein S and Wagner S for TMA construction, optimisation of IHC protocols and performance of sequencing for miRNA.

Statement of parts of the thesis submitted to qualify for the award of another degree

None.

Acknowledgements

Acknowledgement and gratitude is expressed to the School of Dentistry, UQ Centre for Clinical Research and supervisors Professor Camile Farah and Dr Geoff Beadle for supervision of the thesis. Acknowledgement is made for the assistance with technical aspects of the projects particularly to S. Stein, S. Wagner and Y. Emmanuel, staff and other students at the University of Queensland and for the contributions of the co-authors in the respective papers. Dr Suzanne Allen assisted with data collection and the Queensland Cancer Control Advisory Team provided survival information for the project under HREC approval.

Support was provided for the entire PhD project by family and friends.

Funding support was provided by the National Breast Cancer Foundation and the Princess Alexandra Hospital Research Foundation and access to equipment was provided by Queensland Health, Princess Alexandra Hospital.

Keywords

breast cancer, immunohistochemistry, prognosis, artificial neural network, epigenetics.

Australian and New Zealand Standard Research Classifications (ANZSRC)

ANZSRC code: 111201, Oncology and carcinogenesis, 100%

Fields of Research (FoR) Classification

FoR code: 1112, Oncology and carcinogenesis, 100%

Table of Contents

Table of Figures and Tables

Abbreviations

ANN	Artificial neural network
BVI	Blood vessel invasion
CAM	Cell adhesion molecule
CGH	Comparative genomic hybridization
CISH	Chromogenic *in situ* hybridization
CNV	Copy number variation
DCIS	Ductal carcinoma in situ
DFS	Disease-free survival
ER	Oestrogen receptor
FFPE	Formalin fixed paraffin embedded
FISH	Fluorescent *in situ* hybridization
HER2	Human epidermal receptor 2
HIF	Hypoxia inducible factor
HSR	Homogeneously staining region
IHC	Immunohistochemistry
ISH	*in situ* hybridization
LCM	Laser capture microdissection
LEC	Lymphatic endothelial cell
LN(M)	Lymph node (metastasis)
LVD	Lymphatic vessel density
LVI	Lymphovascular invasion
miRNA	micro RNA
MVD	Microvessel density
MR(I)	Magnetic resonance (Imaging)
NO(S)	Nitric oxide synthase
NSLN	Non-sentinel lymph node
OS	Overall survival
PR	Progesterone receptor
QIMR	Queensland Institute of Medical Research
QUT	Queensland University of Technology
RTK(i)	Receptor tyrosine kinase (inhibitor)
RT-PCR	Reverse-transcriptase PCR
SISH	Silver enhanced *in situ* hybridization
SLN	Sentinel lymph node
SLNB	Sentinel lymph node biopsy
TMA	Tissue microarray
TAM	Tumour associated macrophage
UQ	University of Queensland
VEGF(R)	Vascular endothelial growth factor (receptor)

Chapter 1

General Introduction

1.1. Hypothesis

The hypothesis of this study is:

Prognostic and predictive markers defined by immunohistochemistry (IHC) and other histopathologic characteristics of the primary tumour, together with gene expression profiling of the tumour, DNA methylation profiling and microRNA (miRNA) will identify lymph node status in breast cancer patients and allow selection of patients for sentinel lymph node biopsy (SLNB) or axillary lymph node dissection (ALND).

1.2. Aims

- To develop a model for prediction of lymph node involvement in early breast cancer using histologic, phenotypic, expression profiling, DNA methylation, miRNA and genotypic tumour characteristics including IHC markers of cell surface receptor molecule expression, oncogenes, tumour suppression genes and metastasis suppression genes.
- To develop an artificial neural network algorithm to predict lymph node status in breast carcinoma patients from initial core biopsy prior to initial surgery.

1.3. Introduction

The spread of cancer from the point of origin was first termed “metastasis” by Jean Claude Recamier in 1829 (Recamier, 1829). Despite advances in treatment of cancer both in surgical techniques, chemotherapy and targeted therapy, the majority of cancer deaths are due to metastatic disease. The spread of tumour cells that are resistant to therapy occurs from the primary tumour site to distant sites, where they grow, multiply and impact on patient survival.

Most invasive tumours can metastasize, with three pathways of spread to distant organs: direct seeding of surfaces and cavities such as the peritoneal cavity, lymphatic spread and haematogenous spread (Figure 1).

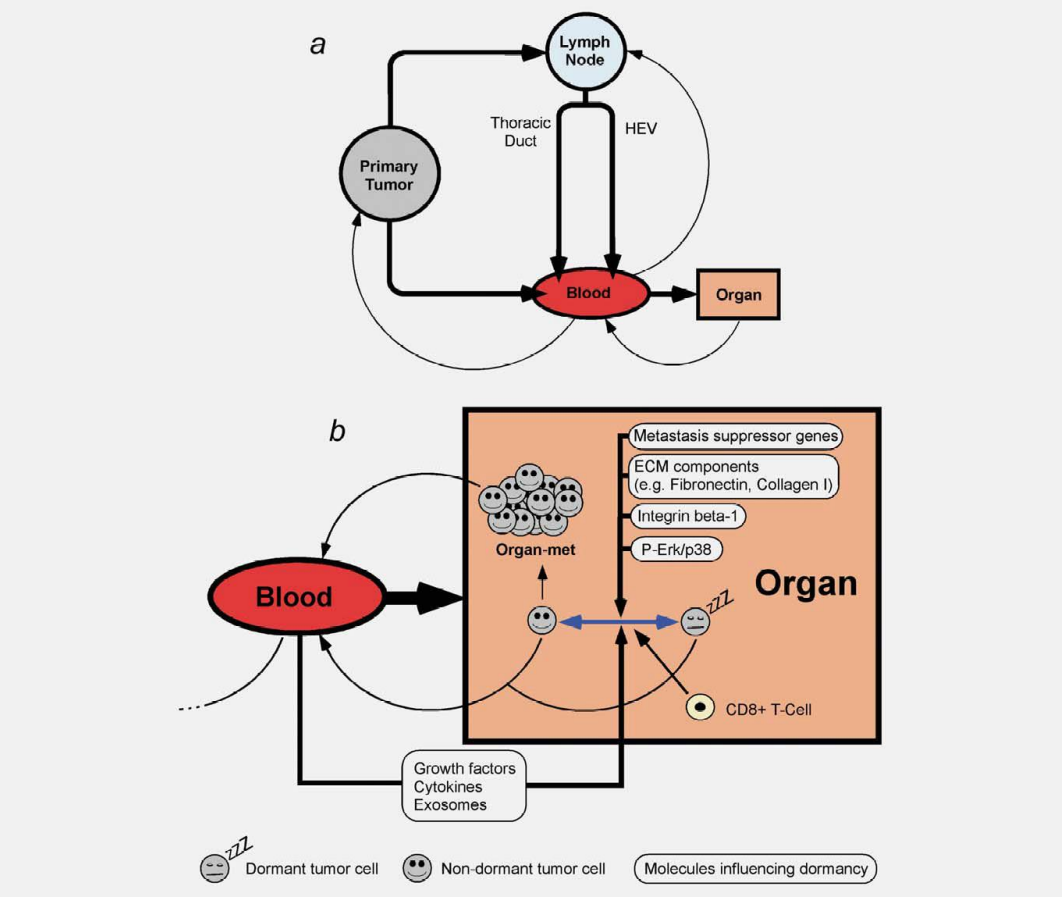


Figure 1: “Routes to metastasis development (a) Schematic diagram showing possible routes tumour cells could conceivably take during their journey through the body. Disseminating tumour cells can leave the primary tumour and enter the blood directly. If they enter the lymphatic system they might lodge in lymph nodes and eventually grow out as lymph node metastasis. Alternatively they could enter the blood stream, for example via the thoracic duct or possibly also via high endothelial venuoles (HEV). Once in the blood stream, disseminating tumour cells that survive might either self-seed the primary tumour, re-enter the lymphatic system or extravasate into organs. Tumour cells shed into the blood by metastases might also participate in self-seeding. (b) Illustration of possible fates of disseminated tumour cells. Once tumour cells have extravasated and invaded vital organs, they either die, enter dormancy or grow out as metastases. A number of factors, including those depicted in the figure, have been implicated in regulating dormancy” (Sleeman et al., 2011).

Dissemination of non-neoplastic cells results in the death of the cell or homing to a particular site e.g. bone marrow precursors. In contrast metastasizing neoplastic cells are not subject to these regulatory influences.

Any primary tumour contains a heterogeneous cell population and for a primary tumour to metastasize requires completion of a complex series of events and interaction with multiple host factors (Fidler, 1991). In 1973 experiments indicated metastasis was a selective process and in 1977 it was identified that malignant tumours were heterogeneous with subpopulations of tumour cells with widely differing metastatic properties (Fidler, 1991). This study showed that the number of lung metastases was proportional to the number of viable tumour cells, however clumps of tumour cells resulted in an increased number of metastatic lesions compared to the injection of an equal number of single tumour cells (Fidler, 1991). New techniques have allowed us to gain some insights into the pathogenesis of metastasis (Tarin, 2006). Multiple steps and multiple genes are required for metastasis to occur: the neoplastic cells must dissociate from the primary tumour, enter the lymphatic system, vascular system or a body cavity, survive during transport, exit the circulation and grow at a distant site (Eccles and Welch, 2007). This process is extremely inefficient and of the approximately 4 million cells per gram of tumour per day entering the circulation less than 0.01% develop into metastatic tumours at other sites (Eccles and Welch, 2007). Many circulating tumour cells are apoptotic or dead and many others are eliminated by shearing forces in the circulation (Sleeman et al., 2011). In addition many circulating tumour cells have the ability to extravasate successfully, but will not form metastases (Sleeman et al., 2011).

Functional genomics has allowed the recognition of tissue-specific metastasis genes for bone and lung sites (Kang, 2006) , but this has not occurred for lymph node metastases. A number of mechanisms have been implicated in this process including selectins, chemokines and their receptors and the concept of premetastatic niches (Sleeman et al., 2011).

All studies have some limitations and this may account for the failure to identify features in primary tumours that relate to lymph node metastases. Immunohistochemistry (IHC) studies are usually small and retrospective. Gene expression profiling studies are generally restricted to fresh or frozen tissue, use an admixture of tumour and stromal cells

and rely on biostatistical processing to obtain any useful information. Animal models may not accurately reflect human tissue features and xenografts usually do not contain supporting human stromal cells. For human studies there is usually little histopathological data (Urquidi and Goodison, 2007) and poor data quality may adversely impact on downstream data.

In particular, contaminating cells may alter the results obtained with expression microarrays. Harrell et al (Harrell et al., 2007) analysed differences between primary tumours and lymph node metastases using tumour sections and laser capture microdissection samples. Less than 1% (30) of genes that varied in expression levels between tumours and the nodal metastases were common to both methods.

Eccles et al (Eccles et al., 2007) recently reviewed lymphatic metastasis in breast cancer and concluded that lymph node metastasis is an important prognostic factor, however, there were a number of unresolved issues and questions. These included:

a) Whether lymph node metastasis is an indicator or governor of metastasis to distant organs.
b) Is lymphangiogenesis required for lymphatic metastasis?
c) Is there a gene signature or other reliable predictor of lymphatic metastasis in primary tumours?
d) What factors are rate limiting for lymphatic metastasis?
e) What is the relative importance of active verses passive mechanisms?
f) How do tumour cells survive in the nodes and escape immune surveillance?
g) Is targeting lymphatic metastasis a viable therapeutic option in cancer?

Adjuvant antiangiogenesis therapy has not improved overall survival in cancer patients and mice models of lymph node metastases has shown no effect of antiangiogenesis therapy on the growth or vascular density of lymph node metastases (Jeong et al., 2015). This suggests that lymphatic metastasis, whilst being a prognostic marker in cancer, may not be a viable therapeutic option for targeted therapy.

Many studies have attempted to define the mechanisms of tumour dissemination in a wide range of tumour types. One of the most widely studied has been breast cancer.

This thesis focuses on lymphatic spread of tumour cells in breast cancer to provide insight into potential patient benefit and treatment therapeutics for breast cancer patients. For

breast cancer, the inability to predict the lymph node status in patients means that a second operative procedure, an axillary lymph node dissection, is currently required if a metastasis is detected in the sentinel lymph node biopsy at the first operation.

1.4. Breast Cancer

Breast cancer is the leading cause of cancer-related deaths in women world-wide (Hinestrosa et al., 2007) and is the second most common cancer in Australian women. More than 13,000 new cases are diagnosed each year in Australia (AIHW and NBCC, 2006), increasing from 11,000 in 2000 (Welfare, 2000),of which 2,200 are in Queensland. Approximately 2,700 deaths from breast cancer occur annually in Australia of which 465 were in Queensland in 2003. The number of hospital separations in Australia for women increased from 15,831 in 1995-96 to 23,598 in 2003-4 (AIHW and NBCC, 2006).

In recent years vast resources have been devoted to attempting to identify and develop biomarkers to help determine which treatments provide the greatest benefit to any given patient with breast cancer. This may be either to select patients for a particular therapy to spare patients from unnecessary or toxic therapy of no benefit. This is considered to be the goal of individualized or personalized medicine. Despite these efforts, biomarker research has generally resulted in ambiguous data with inconclusive findings or insufficient evidence. Molecular profiling may increase the effectiveness of our prognostic profiling, but is still only based on population statistics and is not truly individualized to define treatment for specific patients.

1.5. Breast Pathology

The breast is a modified skin sweat gland that develops into a complex functional structure in the female, but remains as a rudimentary structure in the male.

There has been a gradual increase in the incidence of breast cancer in older women over the last 10 years, but this has probably been due to the introduction of mammographic screening programs. In 1994, the mortality rate from breast cancer started to decline and currently only about 20% of patients are expected to die from the disease.

Ninety-five percent of breast carcinomas are adenocarcinomas with other types such as sarcomas comprising less than 5%. Carcinoma in-situ indicates that the lesion is still

confined within the basement membrane. Invasive carcinoma infiltrates into the adjacent tissue. Larger invasive lesions may cause dimpling of the skin or invade into underlying muscle or chest wall. Lymphatic permeation may occur and lead to thickening of the skin (*peau d'orange*). Approximately 30% of patients will have lymph node metastases at the time of presentation, although many of these will not be palpable.

The aetiology of breast cancer is multifactorial and involves reproductive factors, hormone levels and lifestyle (Ellis et al., 2003). Breast cancer is uncommon in women under the age of 25 except in some familial cases and seventy-seven percent of cases occur in women above the age of 50 (Ellis et al., 2003).

1.6. Somatic genetics of breast cancer

As with other tumours, breast cancer develops through an accumulation of genetic abnormalities by gene mutations, amplification and deletions. No karyotypic hallmarks of breast cancer have been identified (Ellis et al., 2003). Some particular genetic abnormalities have been associated with specific types of breast carcinoma (Thor et al., 2002), but in the majority of cases there are no cytogenetic markers to allow subtyping. Invasive micropapillary carcinoma showed loss of the short arm of chromosome 8 in 16 of 16 cases analysed by comparative genomic hybridization (Thor et al., 2002). Comparisons have also been performed between invasive ductal carcinoma and invasive lobular carcinoma (Turashvili et al., 2007).

1.7. Hallmarks of cancer

There are six hallmarks of cancer that comprise the biological capabilities that define cancer (Hanahan and Weinberg, 2011, Hanahan and Weinberg, 2000). These six characteristics include sustaining proliferative signalling, evading growth suppressors, resisting cell death, enabling replicative immortality, inducing angiogenesis, and activating invasion and metastasis (Figure 2). Genomic instability underlies all of these features. These characteristics have been used to develop drugs that affect the different pathways, however, no well-defined molecular characteristics have been identified to determine a metastatic phenotype.

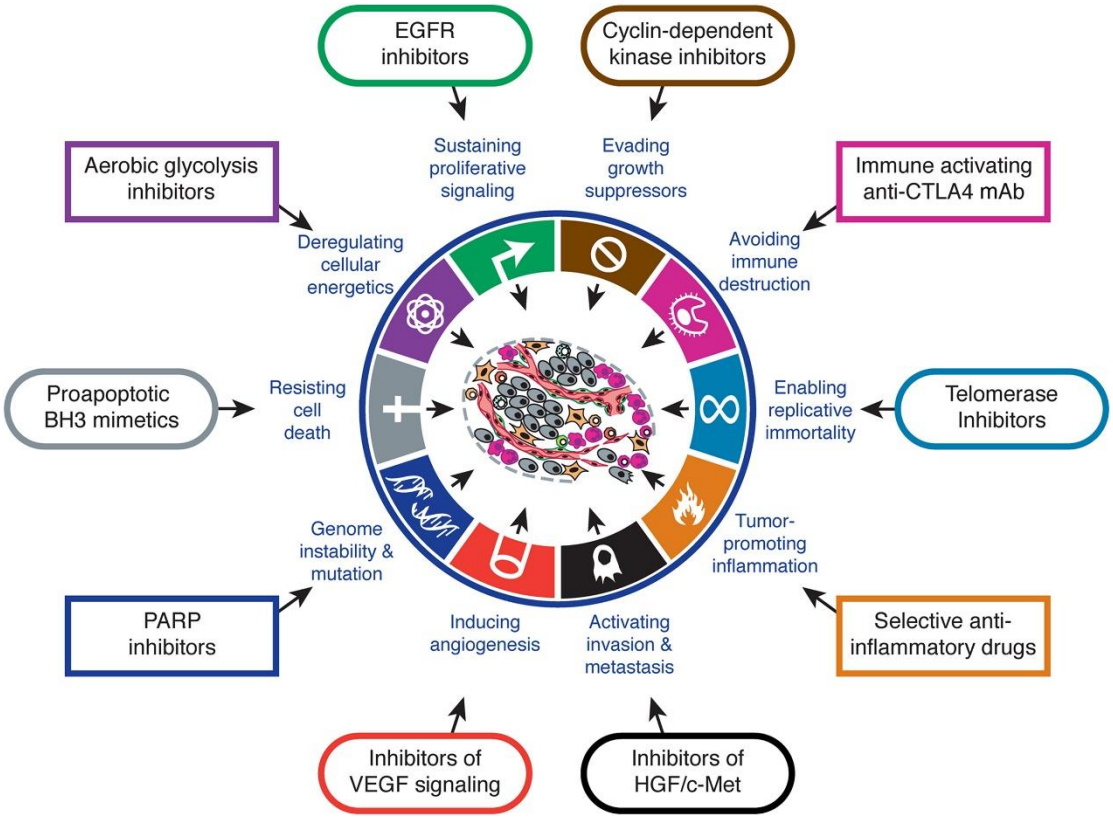


Figure 2: Drugs that interfere with each of the acquired capabilities necessary for tumour growth and progression have been developed and are in clinical trials or in some cases approved for clinical use in treating certain forms of human cancer (Hanahan and Weinberg, 2011).

When the hallmarks of cancer hypothesis was first published in 2000, the mechanisms of metastasis were essentially unknown (Hanahan and Weinberg, 2000). Epithelial tumours have changes in the expression of cell adhesion molecules such as E-cadherin with changes in morphology, undergoing epithelial-to-mesenchymal transition (EMT). In breast carcinoma, lobular carcinomas usually lack E-cadherin, and the lack of expression of E-cadherin by immunohistochemistry staining has been used to differentiate lobular carcinoma from ductal carcinoma. However, the presence of E-cadherin does not exclude metastatic potential as the majority of ductal carcinomas will express this cell adhesion molecule, and a number of these carcinomas will metastasise. Other adhesion molecules associated with cell migration are often upregulated in metastatic tumours (Hanahan and Weinberg, 2011). Whilst extensive research has been performed on metastasis and the process of metastasis, the exact pathways are still largely unknown although EMT may play a role. Kimbung at al (Kimbung et al., 2015) have identified four main groups of genes involved in the metastatic cascade: metastasis initiation genes, metastasis progression genes, metastasis virulence genes and metastasis suppressor genes.

It is also interesting to note, that for haematological metastasis to develop, groups of tumour cells are more likely to form a metastasis compared to individual tumour cells. Recently, quantification of circulating tumour cell clusters (CTC clusters) in blood has shown that whilst they comprise less than 3% of total CTCs, clusters are responsible for more than 50% of metastatic lesions (Bottos and Hynes, 2014). RNA analysis of the CTC clusters identified increased expression of plakoglobin within the clusters (Aceto et al., 2014). The experiments only examined haematogenous dissemination and CTCs, with no data available for lymphatic dissemination.

Inhibitors of c-Met have been developed, but these have had very limited impact on patient care. Abnormalities in the HGF/MET pathway have been reported in multiple tumour types and these changes are associated with tumour stage and prognosis. As this pathway has been associated with invasion and metastasis, the MET pathway has become a potential therapeutic strategy in oncology development in the last two decades. A number of novel therapeutic agents including monoclonal antibodies and small molecule inhibiters, have been tested in patients with different tumour types in clinical studies (Zhang et al., 2011), but the results have generally been disappointing (Van Der Steen et al., 2015). A Phase III clinical trial of Onartuzumab, a MET inhibitor, in NSCLC has been abandoned after the drug failed to show any clinically meaningful efficacy. This drug also failed in a late-stage trial involving triple-negative breast cancers.

1.8. Stem cell hypothesis of breast cancer heterogeneity

Breast carcinomas are heterogeneous with an admixture of cells. The biological mechanisms that result in breast cancer progression and metastatic disease are still poorly understood, but cells can be identified in vascular and lymphatic vessels on histological assessment. The approach used in this thesis attempts to overcome the heterogeneity of the primary tumour by selecting the cells most likely to have the metastatic phenotype. There are two current theories that attempt to explain tumour heterogeneity. These are the cancer stem cell hypothesis and the clonal evolution model. Campbell and Polyak (Campbell and Polyak, 2007) have reviewed these two models. The cancer stem cell hypothesis is that there are a small number of tumour cells with stem cell-like properties. These cells are responsible for initiation, growth and spread of disease (Cariati and Purushotham, 2008). These cells can self-renew indefinitely and survive chemotherapy (Fillmore and Kuperwasser, 2008). The differentiation of these cells leads to the bulk of the

tumour and the cellular heterogeneity. The cancer stem cell hypothesis accounts for metastatic disease by the dissemination of these stem cell-like cancer cells and the resistance to chemotherapy accounts for tumour recurrence after treatment (Cariati and Purushotham, 2008, Stingl and Caldas, 2007, Cobaleda et al., 2008). This concept is not a new one and was first proposed by Cohnheim (Cohnheim, 1875) more than a century ago.

The supporting evidence for the stem cell hypothesis is that the tumour traits could be explained by cells originating from self-replicating stem cells and these cells exhibit pathways associated with stem cells. One of the main arguments to support the stem cell theory is that only a small number of cells are tumourigenic when transplanted into mice. There may be other reasons why this is the case (Kelly et al., 2007, Wicha, 2007) and conclusive proof is still lacking. A large amount of work has been based on leukaemia, but recent studies have identified putative breast cancer stem cells identified by the surface marker CD44. Al-Hajj et al (Al-Hajj et al., 2003, Al-Hajj and Clarke, 2004) showed that CD44+/CD24- phenotype breast cancer cells could form tumours when injected into nude mice.

The second theory is the clonal evolution model which postulates the acquisition of mutations within a tumour and natural selection of the most aggressive clone drives tumour progression. New subpopulations of tumour cells occur resulting in tumour heterogeneity. Any tumour cell may acquire the metastatic or chemoresistant phenotype (Campbell and Polyak, 2007). This theory was first proposed by Nowell (Nowell, 1976) in 1976. Some tumour traits can be explained by this theory such as monoclonality, proliferative capability and heterogeneous cell populations. This would also explain the development of chemotherapy resistance and resistance to other treatments such as tamoxifen and trastuzumab. There is also data showing that primary tumours and metastases are similar genetically (Campbell and Polyak, 2007). Data from DCIS studies also supports this hypothesis (Allred et al., 2008). The most convincing evidence has been obtained using laser capture microdissection. Every major type of tumour contains a number of subpopulations with different abnormalities (Campbell and Polyak, 2007). Similar data has been obtained in breast cancers with matched primary and metastatic tumours (Campbell and Polyak, 2007). Shipitsin et al (Shipitsin et al., 2007) determined molecular profiles on distinct cancer cell populations and concluded that the finding of

genetic differences supported the clonal evolution hypothesis. However, Ailles et al (Ailles and Weissman, 2007) disputed this conclusion and argue that the data from this study support the cancer stem cell hypothesis.

A number of concerns have been raised over the methods used to identify cancer stem cells from solid tumours:

a) Cells are usually isolated from a mixture of fresh tumour cells and the process of isolating the cells may promote selection bias.

b) The cells selected may not be representative of the entire tumour depending on the ratio of tumour cells to stroma and when frozen tissue is used for the isolation of nucleic acid, the cell population is almost always heterogeneous.

c) Studies in mice are not comparable to humans as they usually do not include human stromal elements (Campbell and Polyak, 2007).

In their review of the two different models, Campbell and Polyak (Campbell and Polyak, 2007) propose a combined clonal evolution model that incorporates some features of both hypotheses.

In support of cancer stem cells of cancer stemcell-ness, highly tumorigeneic and metastatic cells have been identified in different tumour types (Li et al., 2015). In this recent study, the small molecular BBI608 inhibited spherogenesis and Stat3-driven transcription, resulting in suppression of metastasis and cancer relapse in xenograft models (Li et al., 2015). The inhibition of metastasis was related to evaluation of direct injection of colorectal cancer cells into the splenic capsule of nude mice, but did not evaluate lymph node metastases in this model (Li et al., 2015).

1.9. Prognostic markers in Breast Cancer

In current clinical practice for breast cancer, the standard prognostic factors that guide adjuvant systemic treatment decisions include tumour size, histologic subtype, nuclear or histologic grade, oestrogen and progesterone receptor status, and axillary lymph node status. A measure of the limitation of these factors in predicting outcome is exemplified by the observation that recurrence rates are approximately 25% in lymph node negative breast cancer patients who will often not receive adjuvant therapy (McGuire and Clark, 1992).

There have been more than 100 variables described to attempt to predict prognosis or response to therapy in breast cancer patients, but only a limited number have entered clinical practice. A predictive factor is defined as a clinical or pathologic feature that determines the likelihood of a response to a particular treatment. A prognostic factor is defined as a clinical or pathologic biomarker that determines patient outcome or survival. A biomarker may be both predictive and prognostic. For a factor to be useful, it must be technically validated, clinically validated and influence clinical decision making. A large proportion of studies have used immunohistochemistry (IHC) to assess the expression of different antigens in breast cancer cells compared to normal tissue. Limitations of these studies include variation in patient selection, different primary antibody clones, different cut-off criteria, interlaboratory variation and reproducibility of the methodology.

In addition, molecular techniques including DNA microarray analysis have indicated a large number of genes are involved in the progression of breast carcinoma. More than 3000 genes have been implicated in distinguishing oestrogen receptor (ERp) positive tumours from ERp-negative tumours (Talisuna et al., 2003).

1.9.1. Grade

Breast cancer is graded according to the Elston-Ellis modified Scarff, Bloom, Richardson grading system also known as the Nottingham modified Bloom and Richardson system (Bloom and Richardson, 1957, Elston and Ellis, 1991). In this grading system, there are three components assessed for assignment of grade to a tumour: Tubule formation, nuclear pleomorphism and mitotic rate. Each component is assigned a number from 1 to 3 and the components are totalled to obtain a value out of 9. Tumours with a score of 5 or less are graded as Grade 1 (G1), 6-7 are Grade 2 (G2) and 8-9 as Grade 3 (G3) tumours. The grade of tumour shows significant correlation with breast cancer specific survival (Burke and Henson, 1997, Nicholas, 1997, Roberti, 1997, Ellis et al., 1992) for patients with lymph node metastases and those without nodal metastases at 5, 10 and 15 years. The evaluation of tubule formation and mitotic rate provide independent prognostic information (Frkovic-Grazio and Bracko, 2002).

Untreated G1 patients have a 5 year survival rate of approximately 95%, whereas those patients with G2 or G3 tumours have 5 year survival rates of approximately 75% and 50%

respectively. Grade is also associated with lymph node metastases and lymph node positivity increases significantly with high-grade tumours (Iwasaki et al., 1998, Ashturkar et al., 2011).

However, there is significant interobserver variability in the evaluation of histological grade by pathologists (Harvey et al., 1992, Robbins et al., 1995, Frierson et al., 1995, Dalton et al., 1994, Dalton et al., 2000) and the usefulness in patient prognosis has been questioned (Younes and Laucirica, 1997, Hayes et al., 2001). Grade 2 tumours comprise approximately 50% of all breast cancers and it has been proposed that refinement of this group may improve the prognostic value of tumour grading (Sotiriou et al., 2006, Ivshina et al., 2006) as Grade 2 tumours usually show the lowest degree of concordance (Rakha et al., 2010a).

1.9.1.1. Molecular Grade

Because of the limitations of histological grading a number of molecular techniques including gene expression profiling (Ivshina et al., 2006, Sotiriou et al., 2006) have been used to attempt improve the current grading system. Sotiriou et al developed a 97-gene expression profile associated with histologic grade and found that the gene expression profile was correlated more strongly with relapse-free survival compared to histologic grade (Sotiriou et al., 2006). Similarly Ma et al (Ma et al., 2003) and Ivshina et al (Ivshina et al., 2006) have also developed histologic grade signatures and the latter enabled the refinement of the G2 tumours into different subgrades with clinical implications (Ivshina et al., 2006). The Genomic Grade Index (Filho et al., 2010, Ignatiadis and Sotiriou, 2008b, Toussaint et al., 2009) has been further developed for both frozen and formalin fixed paraffin embedded tissue (FFPE). Ma et al (Ma et al., 2008) developed a PCR-based five gene molecular grade index (MGI) for applications with FFPE material. The rationale for this approach was that it may be more easily implemented into a diagnostic laboratory. The MGI has been validated in a separate patient cohort (Jerevall et al., 2011) and has also been used with HOXB13:IL17BR for prognosis in lymph node negative patients (Habel et al., 2013) .

1.9.2. Lymph node status

Lymph node metastasis is the main prognostic factor in breast carcinoma.

The mechanism of lymph node metastasis and its involvement in further tumour dissemination is poorly understood. Breast cancer metastasis follows the sequence of invasion into surrounding tissue, penetration into vessels, spread to distant sites and growth at those sites. Breast cancers, in addition to other tumour types, have a predilection to spread to regional lymph nodes (Sleeman, 2000). Once metastasis has commenced, there may be many sites for tumour deposition. Some metastatic deposits may be explainable by purely mechanical forces that trap tumour cells, however there are favoured sites such as the brain and bones that suggest homing mechanisms.

Eccles et al (Eccles et al., 2007) have reviewed lymphatic metastases including mechanisms by which tumour cells may enter lymphatic vessels. Tumours can interact with lymphatic vessels by the production of growth factors and possibly by tumour contribution to the structure of tumour-associated lymphatic vessels (Eccles et al., 2007).

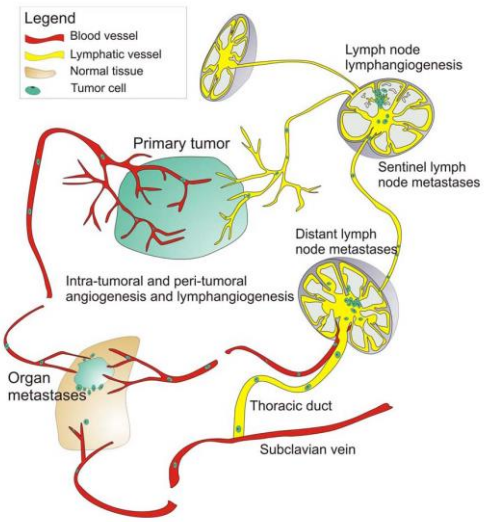


Figure 3: Schematic of tumour-induced lymphangiogenesis at the tumour site and in lymph nodes with metastasis to lymph nodes and distant sites (Mumprecht and Detmar, 2009).

Figure 2 shows the possible paths for metastasis in breast cancer (Mumprecht and Detmar, 2009). It is unknown whether tumour cells need to actively invade into lymphatic vessels or whether this is a passive process (Bockhorn et al., 2007) and also unknown is the mechanism by which tumour cells select either vascular or lymphatic permeation (Wong and Hynes, 2006). These processes may occur in series or parallel and may be determined by distinct molecular mechanisms. If entry into the lymphatic or vascular channels is not the rate-limiting step then other mechanisms must determine the ability of tumour cells to exit the circulation and proliferate at the metastatic sites.

If the cancer stem cell hypothesis is the preferred model for breast cancer, then only cancer stem cells should be capable of giving rise to breast cancer metastases. In support of this theory is the finding that the majority of breast cancer cells identified in the bone marrow are $CD44^{+}/CD24^{-/low}$ (Balic et al., 2006). It has also been demonstrated that these cells express genes involved in pathways associated with cell motility, invasiveness and

chemotaxis (Shipitsin et al., 2007, Liu et al., 2007). However, the $CD44^{+}/CD24^{-/low}$ phenotype was insufficient to identify cells that were able to establish *in vivo* metastases (Sheridan et al., 2006), and only a subpopulation of these cells may be able to perform this function.

Clinical studies suggest that there is no survival difference between patients whose regional lymph nodes have been removed, compared to those who receive no or partial dissection of the nodes (Veronesi et al., 1999, Gervasoni et al., 2000, Thiele and Sleeman, 2006). Therefore this data suggests lymph node metastases are an indicator of systemic disease, rather than having a role in systemic dissemination per se (Eccles et al., 2007), but there are a number of caveats on this interpretation including the role of disturbance of lymphatic drainage due to surgical intervention. Peritumoural lymphatic invasion was associated with distant recurrence and overall survival in a recent study (Arnaout-Alkarain et al., 2007) and this may indicate a role for nodal metastases in seeding distant metastases. However, it may simply indicate that tumour cells disseminate to all sites independently.

Animal models have been used to evaluate distant dissemination of tumour cells in human cancers. In a number of experimental studies an increase in lymphangiogenesis also resulted in increased lymph node and distant metastases (Hirakawa et al., 2007, Krishnan et al., 2003, Skobe et al., 2001), but this is not a consistent finding. Vascular endothelial growth factor (VEGF-C) overexpression in melanomas and fibrosarcomas increased nodal metastases, but did not influence lung metastases (Padera et al., 2002). Conversely the suppression of lymphangiogenesis induction from the tumour resulted in reduced metastases in the regional nodes and in the lung (Krishnan et al., 2003). Similarly the suppression of VEGFR3 and VEGFR2 by antibodies in a breast cancer model (Roberts et al., 2006) resulted in a reduction in regional and distant metastases. The combination of both antibodies inhibited lung and lymph node metastases more effectively than either antibody alone.

Whilst in rats the removal of lymph nodes and primary tumour results in prevention of distant metastases this does not seem to be the same in humans (Ward and Weiss, 1989). Tumour cells may be found in the bone marrow of breast cancer patients who have early stage disease by standard staging systems. Genetic analysis of these cells shows that

they are not identical to lymph node metastases (Schmidt-Kittler et al., 2003, Woelfle et al., 2003), but whether this is due to additional changes or cellular selection is unknown.

Auchincloss published the results of a study in 1963 involving 204 patients between the years of 1951-1953 (Auchincloss, 1963). This article reviewed the survival of 107 patients with lymph node metastases and noted that only those patients with axillary metastases who have three or fewer nodes involved stood much chance of being cured by radical mastectomy.

The cellular and molecular mechanisms of lymphangiogenesis and lymphatic metastasis have been extensively studied and reviewed by Eccles et al (Eccles et al., 2007) and are detailed in Table 1.

Table 1: Lmphangiogenic cytokine expression, lymphangiogenesis and lymphatic metastasis in human breast cancer.

Material	Observation	Reference
107 breast cancers and 22 normal breast samples	HER2 overexpression correlated with VEGF-A, C and D (measuered by IHC) with a trend towards shortened survival	(Yang et al., 2002)
51 breast cancers	Association between HER2 and VEGF-C (measured by IHC) but no correlation between VEGF-C and LNM or LVI	(Hoar et al., 2003)
177 primary invasive breast cancers	High expression of VEGF-A and VEGF-C (IHC) but not VEGF-D was associated with LVD, MVD, LNM and distant metastases	(Mohammed et al., 2007a)
113 and 55 primary breast cancers	High expression VEGF-C correlated with high LVD, LVI, LNM and poor DFS and OS	(Nakamura et al., 2005)
113 primary breast cancers	High expression of nitrotyrosine (NO activity) correlated with VEGF-C expression by IHC and LNM	(Nakamura et al., 2006)
33 primary breast cancers and 7 normal breast samples	No correlation between VEGF-C mRNA and LVI or LNM; VEGF-D levels decreased in tumours and inversely correlated with LNM	(Land et al., 2003)
107 invasive primary breast cancers, LN+	Significant correlations between VEGF-C expression in tumour and TAM (measured by IHC) LVD, LVI and peritumoural inflammatory reaction, CFS and OS	(Sebastian et al., 2006)
217 breast cancers	No correlation between VEGF-C and other prognostic indicators in a diverse patient population	(Yavuz et al., 2005)
80 cases of invasive micropapillary carcinoma of the breast	Positive correlation between VEGF-C and VEGFR-3, higher peritumoural LVD and LVI	(Li et al., 2006)
180 invasive breast cancers	Intratumoural lymphatic vessels in 12% cancers and peritumoural vessels in 94%. High peritumoural LVD correlated with LNM and poor DFS and OS. No correlation with VEGF-C expression.	(Bono et al., 2004)
177 invasive breast cancers	VEGF-C correlated with poor DFS and OS and simulataneous detection of VEGFR-3 increased	(Nakopoulou et al., 2007)

	prognostic value	
29 invasive breast cancers	VEGF-C and VEGF-D expression correlated with LVD. MVD and LVD correlated with LNM	(Choi et al., 2005)
98 breast cancers	VEGF-C protein expression correlated with LVI and poor DFS	(Noguchi et al., 1993)
303 node-negative breast cancers	LVI was the only significant predictor of distant recurrence and OS in a multivariate analysis	(Arnaout-Alkarain et al., 2007)
702 breast cancers, median 8-year follow up	Nanometastases detected in LN in 13% of pN0 patients were a strong indicator of distant metastatic relapse	(Querzoli et al., 2006)
68 invasive breast cancers and 13 normal samples	VEGF-B expression was associated with lymph node status, but not MVD	(Gunningham et al., 2001)
207 invasive breast carcinomas and 15 normal samples	VEGF-D expression was associated with HIF-1α but not lymph node status, LVI, DFS of OS	(Seidman et al., 2001)
105 invasive breast cancers	VEGF-D expression correlated with HER2 expression, LNM and poor DFS and OS	(Nakamura et al., 2003)
75 invasive breast cancers	No evidence of lymphangiogenesis and intratumoural lymphatic vessels (IHC)	(Williams et al., 2003)
23 normal samples, 7 fibrocystic disease, 32 DCIS, 55 invasive breast carcinomas	No evidence of intratumoural lymphangiogenesis, but increased peritumoural LVD compared with non-malignant conditions or DCIS	(Malkas et al., 2006)
13 normal, 11 hyperplastic lesions, 21 DCIS, 40 invasive breast cancers	MVD increased during progression from normal to invasive cancer, but there was little evidence of lymphangiogenesis or intratumoural lymphatic vessels	(Vleugel et al., 2004)
109 invasive breast cancers	Only peritumoural LVI, not BVI correlated with LNM; the former was an independent prognostic marker	(Van den Eynden et al., 2006)
121 ductal carcinomas	LVD, VEGF-C and VEGF-D in primary tumours did not correlate with progression or survival. However, LVD inside LNM correlated with survival. VEGF-D was associated with intralymphatic tumour cells	(van der Schaft et al., 2007)
87 patients with breast cancer and 10-year follow up	VEGF-C expression did not correlate with LNM, LVI, DFS or OS. However expression of CD44 v7-8 correlated with poor DFS and OS	(Watanabe et al., 2005)
123 breast cancer patients	LVI using D2-40 IHC was related to high grade, negative hormone receptor status and worse prognosis	(Marinho et al., 2008)
103 cases of IMPC and 96 cases of IDC	Expression of SDF1 and CXCR4 significantly correlated with number of nodes	(Liu et al., 2009)

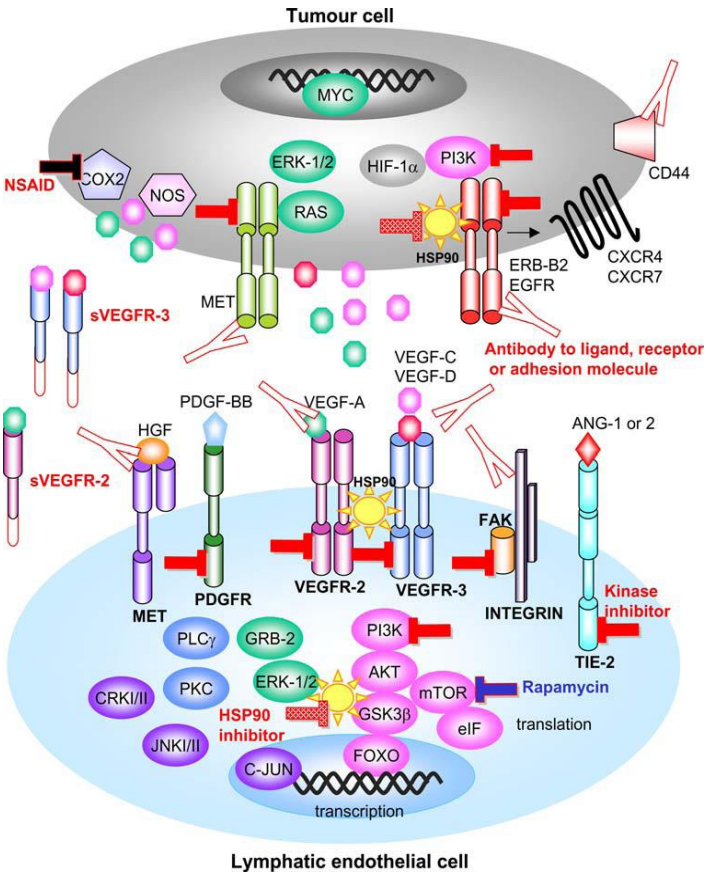


Figure 4: “Signalling pathways in lymphangiogenesis and possible points for therapeutic intervention. Key points for intervention are lymphangiogenic cytokines, their receptors and downstream signalling pathways using antibodies or “decoy” receptors for extracellular targets and small molecule drugs (such as kinase inhibitors and chaperone inhibitors) for intracellular targets. In the latter case, pathways common to both tumour cells and endothelial cells downstream of key RTK (eg. The PI3 kinase pathway) may provide added benefit by inhibiting both tumour and (lymph) endothelial cell functions. Therapeutic benefit may also be obtained by limiting production of lymphangiogenic cytokines in tumour (or host) cells by inhibiting activated oncogenes (eg EGFR, HER2, MET, RAS) or mediators of hypoxic or inflammatory responses (HIF-1α, NOS, COX-2). Finally, processes mediating key cellular interactions between tumour cells and LECs may provide novel, if challenging new targets. These may include cell adhesion, chemotaxis and transendothelial migration potentiated by (for example) integrins, CAMs, chemokines, CD44 etc” (modified from Eccles et al (Eccles et al., 2007)).

The various factors associated with lymphangiogenesis are illustrated in **Error! Reference source not found.**3. Mumprecht and etmar (Mumprecht and Detmar, 2009) have also recently reviewed lymphangiogenesis and cancer metastasis. This article reviewed the evidence for tumour-induced lymphangiogenesis as a predictive indicator for lymph node metastases and possible targets for prevention of nodal metastasis. However, the evidence for correlation between angiogenic cytokines and lymphatic vessel density or lymph node metastases is circumstantial and the mechanisms in clinical metastasis remain to be established.

The presence or absence of lymph node metastases has been included in the TNM classification system (Singletary et al., 2003, Singletary and Greene, 2003) with modifications to the original scheme based on the method of detection of lymph node metastases and the size of the deposits. There is still debate about the prognostic significance of small tumour deposits (Dowlatshahi et al., 1997b, Mansi et al., 1999, Dowlatshahi et al., 1997a, Steinhoff, 1999, Querzoli et al., 2006, Mittendorf and Hunt, 2007, Chen et al., 2007, Grabau et al., 2007, Marinho et al., 2006, Maibenco et al., 2006, Herbert et al., 2007). These were first documented by Pickren in 1961 (Pickren, 1961) with a review of the pathology of the cases included in Auchincloss’s study (Auchincloss, 1963). Serial sections were taken from 51 patients in which previous routine sectioning had not shown nodal metastases. Additional occult metastases were found in 22%. The 5

year survival rate of 91% was similar to that for patients with no occult metastases (89%). Herbert et al (Herbert et al., 2007) in a study assessing isolated tumour cells in lymph nodes showed no difference in patient survival or disease free survival over 2.5 years for 16 patients, similar to this initial study. However, Querzoli et al (Querzoli et al., 2006) demonstrated that nanometastases are a strong risk factor for disease-free survival and metastatic relapse in breast cancer patients.

1.9.3. Sentinel lymph node biopsy (SLNB)

Early lymph node involvement gave rise to the concept of sentinel lymph node biopsy. The sentinel lymph node is the first lymph node in a lymphatic drainage area that receives lymph flow from the primary tumour. Therefore it should be the first lymph node to develop metastatic disease.

For small invasive breast carcinoma, sentinel lymph node biopsy is rapidly becoming the standard of care and is replacing axillary lymph node dissection as the preferred means to evaluate nodal status in patients (Ozmen and Cabioglu, 2006, Bedrosian et al., 2000, Moffat, 2001, Borgstein et al., 1997, Bower, 1997, Cohen et al., 2000, Cox et al., 1998, Dixon, 1998, Donohoe, 2001, Jani et al., 2003, Kollias et al., 1999, Krag et al., 1998, Krag et al., 1993, Kunkler et al., 1999, Masters et al., 1998, Nieweg et al., 1997, Rovere and Bird, 1998, Rozenberg et al., 1999, Schillaci and Scopinaro, 1997, Schrenk et al., 2000, Schwartz et al., 2002, Schwartz and Meltzer, 2003, Singletary and Greene, 2003, Thornton, 1999, Tsang, 1999, Ung and Wetzig, 1999, Veronesi, 1997, Veronesi et al., 1997, Wong, 2002, Mansel et al., 2006). This appears to result in decreased morbidity for patients with reduced relative risks of lymphoedema and sensory loss, as reported in the ALMANAC trial (Mansel et al., 2006). Of concern however is the fact that if the sentinel lymph node is positive, the patient must undergo a full axillary lymph node dissection, resulting in a second procedure.

Analysis of the sentinel node should also indicate whether more widely disseminated disease is present, however about 20% of patients with node-negative disease will go on to develop distant metastases (Rosen et al., 1981). Whilst this data suggests that about 20% of systemic metastases are derived from tumour cells that bypass the lymphatic route (Fisher et al., 2004), this may depend on the method of detection of metastases in the lymph node and the initial lymph node status may have been a false negative result.

For many tumour types there may be little survival advantage to removal of the lymph nodes, but it is also possible that in some cases, the lymph nodes act as reservoirs to seed distant metastases (Sleeman et al., 2011).

1.9.4. miRNA and metastasis

MicroRNAs (miRNA) are non-coding regulatory RNAs that were originally discovered because of the role in controlling the timing of *C. elegans* larval development. More than 900 human minas have been identified and these are catalogued and annotate in a database (miRBase)(2011, Kozomara and Griffiths-Jones, Griffiths-Jones, Griffiths-Jones et al., 2008, Griffiths-Jones, 2006, Griffiths-Jones et al., 2006).

The primary transcripts of miRNA (pri-miRNA) are transcribed by RNA polymerase II before capping, polyadenylation and maturation of a hairpin loop structure by Drosha into pre-miRNA (Hurst et al., 2009b). The stem-loop pre-miRNA is exported out of the nucleus and further processed in the cytoplasm by Dicer to yield a functional single-stranded mature RNA construct of 20 to 22 bases (Sayed and Abdellatif, 2011). This process is illustrated in Figure 5.

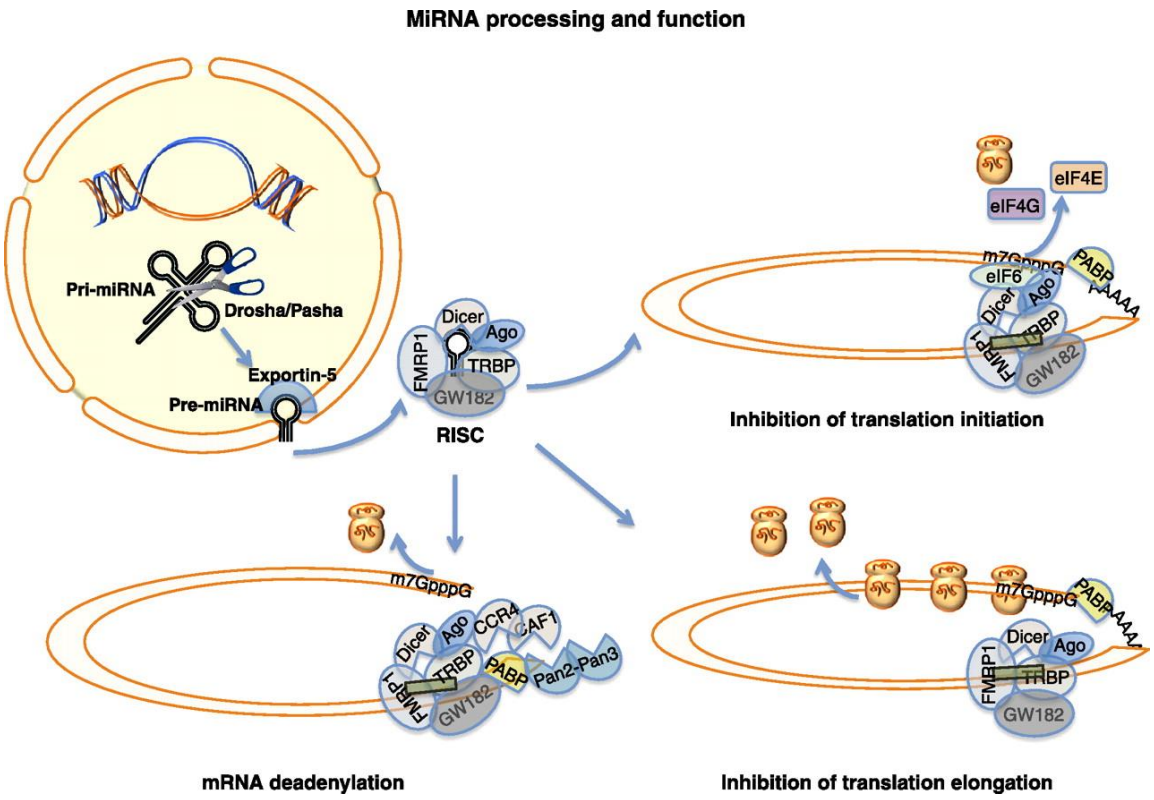


Figure 5: “An illustration of the steps and molecules involved in the processing and function of minaRNAs. The first processing step occurs in the nucleus, where pri-miRNA is cleaved via the enzymes Drosha and Pasha. The pre-miRNA hairpin structure is then exported to the cytoplasm via exportin-5. In the cytoplasm it assembles in RISC which includes Dicer, argonaute (Ago) protein, PW182 a P-body protein, the human immunodeficiency virus transactivating response RNA-binding protein (TRBP), and fragile X mental retardation protein (FMRP1). In RISC, the loop of the pre-minas is cleaved and the mature strand is used to direct the complex to its target site in the 3=-UTR of an mRNA. RISC is involved in 3 different functions: inhibition of translation initiation, inhibition of translation elongation, or mRNA deadenylation, which would result in mRNA degradation. These functions require additional recruitment of eukaryotic initiation factor 6 (eIF6). Inhibition of translation initiation is a result of disengagement of the 80S ribosomal complex in a cap (m7GpppG)-dependent manner that is partly attributed to the recruitment of the anti-association factor eIF6 by the RISC complex, or competition of Ago2 with eukaryotic initiation factor 4E (eIF4E) or eIF4G. Deadenylation occurs via interaction of Ago2 with chromatin assembly factor 1 (CAF1) and, thus, recruitment of the CCR4-NOT protein complex that harbours both chemokine (C-C motif) receptor (CCR4) and CAF1 deadenylase enzymes, in addition to a second deadenylation step involving the PABP-dependent poly(A) nuclease 2 (Pan2)-Pan3 deadenylase complex” (Sayed and Abdellatif, 2011).

The interaction of a miRNA with its target RNA can result in mRNA cleavage or translational repression (Gotte, 2010). A number of miRNAs have been implicated in critical steps in the metastasis pathway in solid tumours (Hurst et al., 2009b), including epithelial-mesenchymal transition, apoptosis and angiogenesis (Hurst et al., 2009b). These critical steps are outlined in Figure 6.

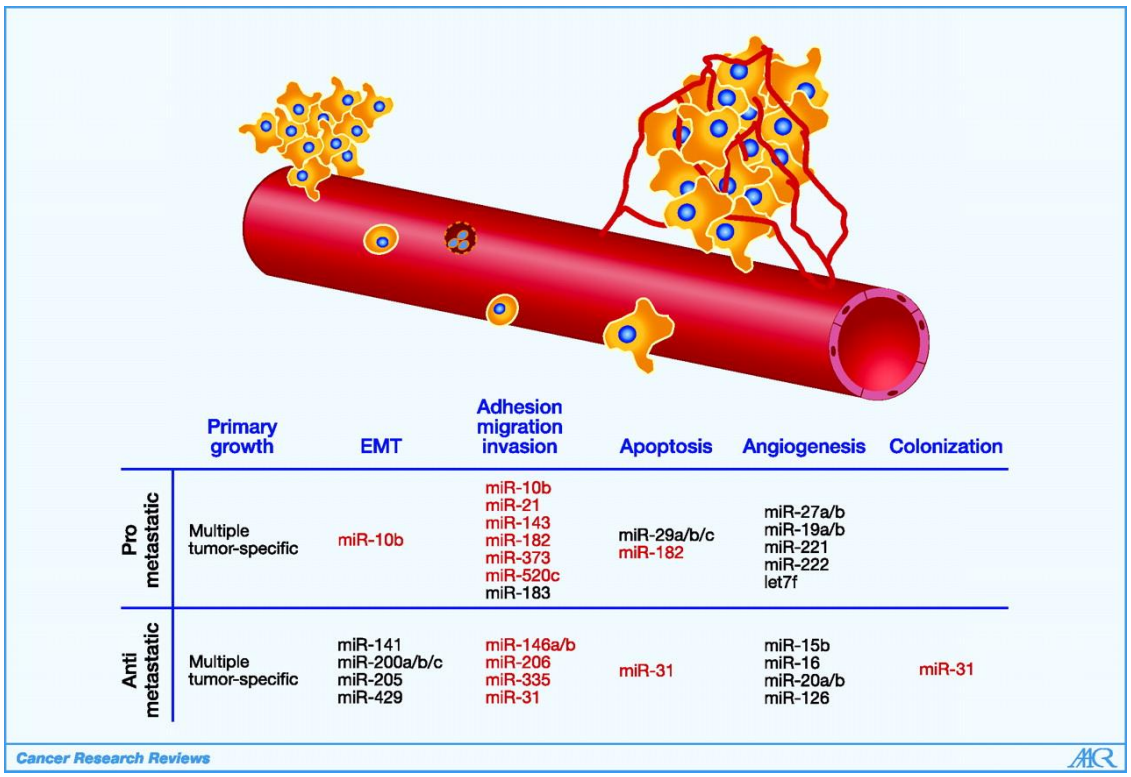


Figure 6: Critical steps in metastasis altered by metastamir.
Pro- and antimetastatic metastamir are listed with the steps in the metastatic cascade of which they affect. The metastamir that have been functionally tested for metastasis in vivo are highlighted in red (Hurst et al., 2009b).

Profiles of miRNA expression have been specifically documented for breast cancer with correlations between miRNA expression and invasion, recurrence, metastasis and survival (Shi et al., 2010). A more detailed breakdown of miRNAs associated with breast cancer metastasis is illustrated in Figure 7.

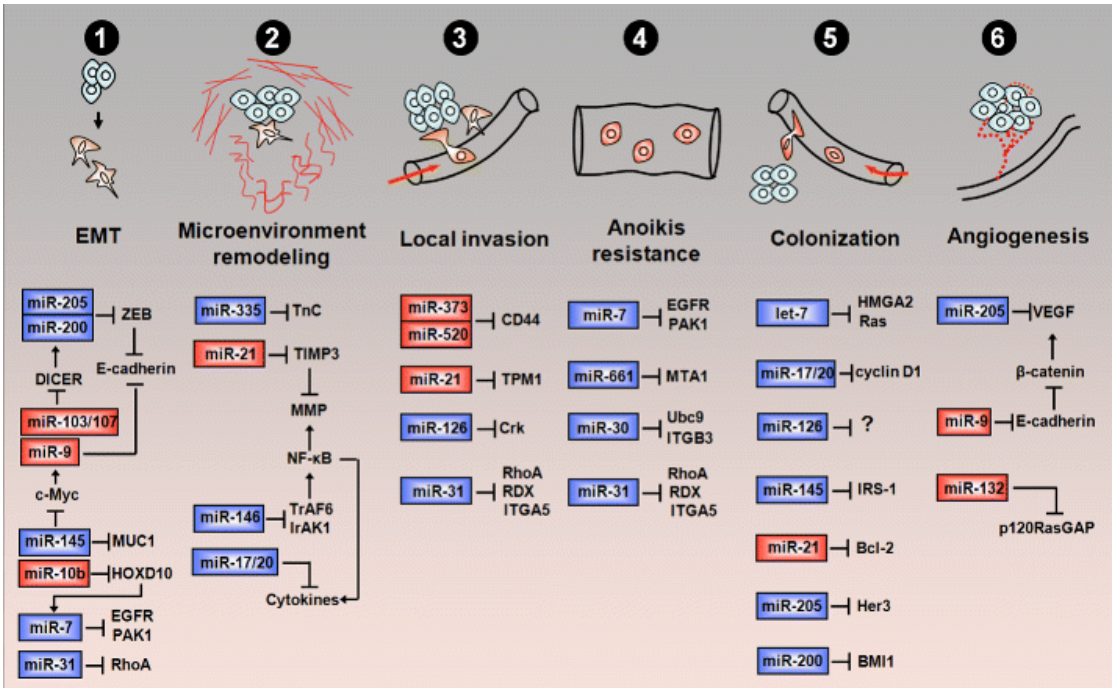


Figure 7: “miRNAs control breast cancer metastasis by influencing multiple metastatic steps. (*1*) miRNAs involved in regulation of EMT: miR-205 and miR-200, may dominate EMT and metastasis by inhibiting the expression of transcriptional repressors of E-cadherin ZEB1 and ZEB2. miR-103/107 modulates the expression of DICER, causing an extensive downregulation of mature miRNAs, including miR-200. miR-9 could inhibit the expression of E-cadherin by binding to the 3′ UTR of E-cadherin, whereas the expression of miR-9 in breast cancer cells could be activated by Myc. Enhanced invasion and metastatic potential mediated by c-Myc and MUC1 could be blocked by miR-145. miR-10b may inhibit HOXD10, leading to an increased invasiveness and metastasis of tumor cells. As a downstream target gene of HOXD10, miR-7 was suggested to negatively regulate the metastatic capability of breast cancer cells by inhibiting EGFR and PAK1 expression. miR-31 regulates a cohort of EMT related genes including RhoA. (*2*) miRNAs involved in remodeling of tumor microenvironment: miR-335 may modify tumor microenvironment by directly influencing the expression of TnC that can reduce cell-ECM interaction. miR-21 affects tumor invasion and metastasis by targeting an inhibitor of MMPs, TIMP3. miR-146 negatively regulated NF-κB activation, which promotes the activities of MMPs by repressing TrAF6 and IrAK1 in the Toll-like receptor and IL-1 signaling pathways, leading to the downregulation of the NF-κB target genes IL-8, IL-6, and MMP-9. miR-17/20 may regulate tumor microenvironment by inhibiting the secretion of the cytokines required for metastasis. (*3*) miRNAs associated with local invasion: miR-373 and miR-520c exerted the pro-invasive functional effects partially by preventing the translation of CD44. miR-21 may affect tumor invasion by targeting TPM1. miR-126 targets CrK, which was involved in the events required for tumor cell migration. miR-31 regulated a cohort of invasion/metastasis-related genes, including ITGA5, RDX, and RhoA. (*4*) miRNAs associated with anoikis resistance: miR-7 may regulate EGFR and PAK1 expression, influencing tumor cell survival. miR-661 may repress MTA1 and regulate anchorage-independent growth of breast cancer cells. miR-30 may inhibit anoikis resistance and metastasis by targeting Ubc9 and ITGB3. miR-31 may inhibit apoptosis and anoikis by affecting the expression of RhoA, RDX, and ITGA5. (*5*) miRNAs involved in tumor cell colonization: let-7 can inhibit proliferation and mammosphere formation of tumor cells by interfering with the expression of HMGA2 and Ras. miR-17/20 may inhibit breast cancer cell proliferation by negatively regulating cyclin D1. Mir-126 may suppress cell cycle progression. miR-145, miR-21, miR-205, and miR-200 inhibit tumor cell colonization by regulating IRS-1, Bcl-2, Her3, and BMI1, respectively. (*6*) miRNAs involved in angiogenesis: miR-205 may inhibit the suppression of VEGF-A expression by binding to the putative miR-205 site in the 3′ UTR of VEGF-A. Downregulation of E-cadherin by miR-9 allows liberation of β-catenin, which then translocates to the nucleus and activates the expression of VEGF, resulting in increased tumor angiogenesis. Upregulation of miR-132 in endothelial cells enhanced neovascularization in breast tumors by inhibiting endothelial p120 Ras GAP expression. miRNAs in *red rectangles* play stimulatory roles in breast cancer metastasis; miRNAs in *blue* may inhibit breast cancer metastasis” (Shi et al., 2010).

Specific miRNAs have been associated with multiple steps in invasion and metastasis: Tumour microenvironment modification (miR-335, miR-17/20, miR-146), Breast cancer stem cell phenotype formation (let-7, miR-200, MiR-30), Epithelial-mesenchymal transition (miR-200, miR-10b, miR-145, miR-103/107, miR-9), local invasion (miR-21, miR-126, miR-373, miR-520), survival in vasculature (miR-7, miR-661, miR-17/20) and proliferation at distant sites (miR-200, let-7) (Shi et al., 2010, Sayed and Abdellatif, 2011).

As indicated in Figure 6, the downregulation of miR-200c has been implicated in epithelial-mesenchymal transition by releasing ZEB1 and suppressing E-cadherin (Hurteau et al., 2007) and miR-200 family members are specifically lost in metastatic breast cancer cells that lack E-cadherin (Sayed and Abdellatif, 2011). Other miRNAs have been identified through profiling of human breast cancer cell lines with various metastatic potential and this approach identified miR-335 and miR-206 as being reduced during metastasis (Tavazoie et al., 2008). Another miRNA associated with metastasis is miR-31. It has been reported to be downregulated ~ 4 –fold in tumour cells compared to normal cells with a reduction of ~100-fold in metastatic cells (Valastyan and Weinberg, 2010, Valastyan et al., 2009b). There is reduced expression of miR-31 in breast, prostate, ovarian and gastric carcinomas, but there is reported up-regulation in colorectal, liver and head-and-neck cancers (Valastyan and Weinberg, 2010). Other miRNAs have been found to promote metastasis including miR-10b (Ma et al., 2007a), miR-373 and miR-520c (Huang et al., 2008). miR-10b is induced by Twist (Yang et al., 2004) and directly targets HoxD10 which is involved in metastasis (Ma et al., 2007a).

However, despite the promising results from the recent publications on miRNAs,there is still a lack of data in clinical samples and the impact of dysregulation of miRNAs on pathological processes including metastatic dissemination.

1.9.5. Molecular Classification of Breast Cancer

Molecular techniques have been applied to breast cancer to attempt to improve the prognostic and predictive value of the current classification systems because of the limitations. Gene expression profiling was used (Hu et al., 2006, Weigelt et al., 2005a, Perou et al., 2000) to molecularly classify breast cancer into five groups: luminal (A & B), HER2-enriched, basal-like and normal-like. This classification provided prognostic

information and predicted response to chemotherapy (Viale, 2012) but had significant practical limitations. The original assay was based on fresh frozen tissue and was not applicable to formalin fixed paraffin embedded material. This meant that this molecular classification was not able to be implemented into routine diagnostic practice. In an attempt to overcome the limitation of the assay, alternate methodologies were developed as surrogate markers to obtain similar information that could be used to improve the classification of breast cancer. Immunohistochemistry (IHC) for oestrogen receptor (ER), progesterone receptor (PR), HER2 and Ki-67 were used to approximate the molecular classification of breast cancer. ER, PR have been in use for more than 20 years as the standard of care to select patients for hormonal therapy as part of a treatment regime. HER2 IHC in Australia is relatively poorly performed and the standard of care for HER2 assessment is in-situ hybridisation (ISH) (Torlakovic et al., 2015, Torlakovic et al., 2014, Francis et al., 2007a). Ki-67 was implemented as a proliferation marker and these four biomarkers were used to mimic the molecular classification derived from gene expression profiling (ER+/PR+/HER2-/Ki-67low=luminal A; ER+/PR+/HER2-/Ki-67high=luminal B; HER2+ = HER2; ER-/PR-/HER2- = basal). However, the classes defined by gene expression profiling and those defined by IHC do not completely overlap (Viale, 2012). More recently, a gene expression approach using quantitative real-time reverse transcription-polymerase chain reaction (qRT-PCR) for 50 genes (PAM50) has been developed that works on FFPE tissue (Parker et al., 2009), however, this has not entered routine clinical practice as it requires specialised equipment that is not routinely available in a diagnostic pathology laboratory.

1.9.6. Epigenetics and metastasis

Epigenetics has emerged in the last decade as potential biomarkers in cancer. Epigenetic modifications do not change the DNA coding, and include changes such as aberrant gene promoter DNA methylation and histone protein modifications. DNA hypermethylation and DNA hypomethylation both occur in a variety of cancers (Ehrlich, 2002). For epigenetic studies, the degree of methylation is usually compared to normal tissues. In cancer, hypermethylation of the genome usually occurs in CpG islands in gene regions. This is in contrast to the areas of hypomethylation which occur in cancer and are seen in heterochromatic DNA repeats, dispersed retrotransposons and endogenous retroviral elements (Ehrlich, 2002).

Hypermethylation of CpG islands in the promoter region of genes leads to epigenetic inactivation of genes associated with cancer regulation and this may account for tumour development or disease progression. Since metastatic disease and lymphatic metastases occur as part of cancer progression, a possible cause of metastasis may be due to epigenetic modifications, rather than DNA somatic mutations.

Melchers et al (Melchers et al., 2015) looked at 28 methylation markers in oral and oropharyngeal squamous cell carcinoma and the association with lymph node metastases. In this study two hypermethylated genes, *MGMT* and *DAPK1* were associated with lymph node status. *MGMT* methylation was associated with a negative lymph node status and tumours also showed reduced protein expression with immunohistochemistry. *DAPK1* methylation was associated with a positive lymph node status, but hypermethylation did not associate with protein expression on immunohistochemistry (Melchers et al., 2015).

In colorectal cancer, DNA hypermethylation of the promoter of *GCNT2* is associated with aberrant expression of *GCNT2* and hypomethylation of *GCNT2* variant 2 has been shown to be associated with lymph node metastasis in primary colorectal carcinoma (Nakamura et al., 2015). A CpG island methylator phenotype (CIMP) has also been identified in colorectal carcinoma where hypermethylation of a number of genes is associated with malignancy (Tahara et al., 2014, Toyota et al., 1999). Hypermethylation of CpG islands and a correlation with lymph node metastases has also been examined in melanoma and breast cancer. In melanoma, Cadherin-11 (*CDH11*) was shown to be hypermethylated in the cell Ines derived from lymph node metastases from patient tumours, but was unmethylated in the cell lines derived from the primary tumour (Carmona et al., 2012). Other groups have also identified aberrant methylation patterns in melanoma and correlated these patterns with lymph node metastases (Huynh and Hoon, 2012).

In breast carcinoma, *ID4* hypermethylation has been reported to occur more frequently in lymph node-positive tumours, compared to lymph node-negative patients (67% vs. 19 %, respectively) and multivariate logistic regression showed that *ID4* methylation was a significant risk factor for nodal metastasis (Huynh and Hoon, 2012). In a separate analysis of 151 breast cancers for the methylation status of six genes, hypermethylation of *GSTP1* and/or *RARβ2* was significantly associated with patients who had macroscopic disease in a sentinel lymph node compared with those with microscopic or no sentinel lymph node metastases (Huynh and Hoon, 2012). In this study, hypermethylation of *CDH1* was reported to be higher in the sentinel lymph node metastasis compared to the primary tumour, but this was not confirmed in a separate study (Huynh and Hoon, 2012). Whilst

these studies indicate that epigenetic events may play a role in metastasis in the early stages of solid tumours including breast cancer, the genes involved have not been well defined. It is also possible that the changes identified in the lymph node metastases are due to the microenvironment of the lymph node itself and not the causation of the nodal metastasis.

1.10. Prediction of lymph node status in breast cancer

1.10.1. Prediction of non-sentinel lymph node (NSLN) metastases following SLNB

Clinicopathological models have been developed to predict the risk of patients having additional lymph node metastases with a positive sentinel node (Cserni, 2007, Cserni et al., 2007, Chagpar et al., 2006, Coutant et al., 2008, Lambert et al., 2007, Bevilacqua et al., 2007, Alran et al., 2007, Pal et al., 2007, Konecny et al., 2003, Kapur et al., 2007, Ponzone et al., 2007, Dauphine et al., 2007, Bassi et al., 2006, Smidt et al., 2007, Werling et al., 2003, Degnim et al., 2005, Carmon et al., 2006, Van Zee et al., 2003, van Deurzen et al., 2007). The MD Anderson Cancer Centre model developed a mathematical scoring system to estimate the risk of a patient having a non-SLN metastasis (Werling et al., 2003) and this has been further validated using an external database (Lambert et al., 2007). A separate model was developed by the Memorial Sloan-Kettering Cancer Center (MSKCC) (Van Zee et al., 2003). All of these models however have used relatively limited pathological information. Dauphine et al (Dauphine et al., 2007) evaluated three scoring systems and whilst these provide additional information, their predictability remains less than optimal. These current studies indicate that axillary lymph node dissection is still required if the sentinel node is positive, as involvement of non-sentinel lymph nodes cannot be predicted with any degree of accuracy and non-sentinel nodes may be involved in up to 60% of cases.

1.10.2. Prediction of axillary lymph node involvement

The prediction of axillary nodal involvement from features of the primary tumour characteristics was described in 1967 (Kouchoukos et al., 1967) and larger evaluations have occurred since (Gann et al., 1999). Kouchoukos et al (Kouchoukos et al., 1967) assessed 432 breast cancers and classified them into 4 subtypes based on histological appearance. The histological appearance, size of the primary tumour and the type of margins were evaluated to assess the usefulness in predicting axillary lymph node

involvement. The accuracy of predicting lymph node involvement for invasive tumours did not exceed 70%.

The presence or absence of lymph node metastases is the most powerful predictor of disease outcome (McGuire, 1987, Foster, 1996, Weigelt et al., 2005c). Gann et al (Gann et al., 1999) commented in 1999 that the difficulty of using predictive models to avert axillary lymph node dissection should not be underestimated. In this study, lymph node status was assessed on 18,025 patients. A number of factors were associated with lymph node status including tumour size and grade, older age, white race, an inner quadrant tumour site, various histologic patterns, negative ERp and PR status and DNA content. They also noted that the benefits of using models to predict lymph node status include lower surgical morbidity and cost; however, a patient who is affected by an incorrect decision will either forgo a significant survival advantage or unnecessarily suffer the morbidity of chemotherapy.

Lymphovascular invasion is associated with lymph node status (Guarnieri et al., 2001, Kurosumi et al., 2001, Anan et al., 2000, Promish, 1999, Olivotto et al., 1998) but, as there is not an absolute correlation, is insufficient for accurate prediction. A number of biomarkers, detected using different methodologies, have been added to the morphological features in an effort to improve the predictive utility of any models. Potential biomarkers and methods include cell surface markers detected by flow cytometry (Menard et al., 1995, Mattfeldt et al., 2004, Mannweiler et al., 2002, Ravdin et al., 1994, Friedman and Freedman, 1994, Ahlgren et al., 1994), more accurate detection of lymphovascular invasion using IHC for identification of lymphatic and vascular channels (Braun et al., 2007, Van den Eynden et al., 2006, Mohammed et al., 2007b), MR-determined metabolic phenotype (Bathen et al., 2007, Sitter et al., 2006), MR imaging of lymph nodes(Mussurakis et al., 1997), free circulating tumour DNA with mutation analysis (Umetani et al., 2006), aberrant DNA methylation patterns (Umetani et al., 2005), IHC biomarkers on the primary tumour (Skobe et al., 2001, Bono et al., 2004, Gunningham et al., 2001, Li et al., 2006, Nakamura et al., 2005, Nakamura et al., 2003, Nakamura et al., 2006, Noguchi et al., 1993)and proteomics to identify metastasis associated factors (Kreunin et al., 2007, Nakagawa et al., 2006).

There have been a limited number of microarray studies performed to attempt to predict nodal status by the identification of genes that are differentially expressed in lymph node metastases compared to primary tumours (Weigelt et al., 2005c, Lee et al., 2003, Montel et al., 2005, Feng et al., 2006, Hao et al., 2004).

Two of these studies used xenograft models of breast cancer (Lee et al., 2003, Montel et al., 2005). Using MDA MB 435 breast carcinoma cells, one study compared primary tumour cells from this cell line with cells derived from axillary nodal metastases (Lee et al., 2003). Thirty-six up-regulated genes and 17 down-regulated genes were identified in the primary tumours with increased metastatic potential. These genes included CD73, integrin α1, integrin β5, HIF1α and genes associated with tumour-matrix interactions. Montel et al (Montel et al., 2005) used the same model and identified 50 differentially expressed genes including the RAS oncogene family of genes and MMP14. None of the genes generally accepted as mediators of lymph node metastases were identified in these studies. These studies used different platforms, different types of samples and different sites of sampling, all of which may have had an effect on the results.

Four studies used patient derived samples. Twenty-six primary tumours and the matched lymph node pairs were evaluated in one study (Feng et al., 2006). 21 up-regulated and 58 down-regulated genes were identified in the nodal metastases including genes associated with collagen, basement membrane-interacting molecules and MMPs.

A second study evaluated 15 primary tumours and the lymph node pairs (Hao et al., 2004). 71 up-regulated and 67 down-regulated genes were identified including those for L-selectin, *EphA2*, ephrin B3, *IGFBP5* and zinc proteins. Interestingly, *MMP2* and fibronectin 1 gene expression levels did not correlate with protein expression as determined by IHC.

A third study of 15 primary tumours and matched lymph node pairs using expression microarrays found no common gene that was differentially regulated (Weigelt et al., 2005c). The pattern of gene expression was different across samples and a classifier could not be developed to predict the nodal status of 295 primary tumours.

A fourth study evaluated 89 primary tumours, 70 with lymph node metastases and 17 without to identify a metagene signature associated with lymph node status and

recurrence in 90% of cases (Huang et al., 2003). This study by Huang et al (Huang et al., 2003) did identify patterns of expression associated with nodal status in 89 Taipei patients, however when this was evaluated in a second patient dataset, lymph node status could only be accurately predicted in 50% of cases (Weigelt et al., 2005c). The data from this study suggested that lymph node metastases occurs independently of distant metastases (Weigelt et al., 2005c).

A recent study also evaluated lymph node metastases and their matched primary breast cancers using gene expression profiling (Suzuki and Tarin, 2007). In this study it was observed that the metastases had very similar expression signatures to their parent tumours and a small number of genes were consistently differentially expressed. The conclusion from this study was that the data is most consistent with the coexistence of a number of cell clones in the primary tumour, each possessing randomly different parts of the gene expression pattern required to accomplish metastasis, co-evolving to metastatic status, but collectively possessing an “average” signature typical of distant metastases. This reinforces the clinical observation that in most patients with small tumours, metastasis is not inevitable.

Ellsworth et al (Ellsworth et al., 2011) performed gene expression profiling on 76 breast cancer patients. Statistical analysis of the expression profile of tumours with and without lymph node metastases showed significant differences in the expression level of 15 probes corresponding to 11 genes (Downregulated genes: *ABCC8*, *BATF*, *IGFBP6*, *MRPL40*, *SLC27A2*; Upregulated genes: *AURKA*, *KIAA1609*, *KIF23*, *PLCB1*, *RPL13*, *TCP1*, *TGFA*). The number of genes differentially expressed did not differ significantly and hierarchical clustering analysis was able to correctly classify 90% of the lymph node-negative tumours (37/41), but only 66% of the lymph node-positive tumours (23/35) (Ellsworth et al., 2011). Possible reasons identified by the authors for the failure to identify a gene signature include the study design, biological properties of the primary breast cancers, molecular heterogeneity, the microenvironment and inherent host susceptibility (Ellsworth et al., 2011).

Gene expression profiling has been used to predict metastasis from low-risk breast cancer patients (Thomassen et al., 2007) and a 14-gene predictor has been described with a sensitivity of 83% and a specificity of 76% (Ma et al., 2007b). Urquidi and Goodson

(Urquidi and Goodison, 2007) combined genomic signatures of breast cancer metastasis from tissue based studies and animal models. Whilst this was not specific for lymph node metastases a number of points were noted in this study. Future directions to improve data quality included the provision of more accurate histopathological data and microdissection of tumour cell populations with separation from stromal cells. It was also noted in this article that genomic approaches are unlikely to be adequate as a sole prognostic and predictive platform in breast cancer and that validation of array data should be performed at the protein level wherever possible.

Using patient tissues it has been possible to develop a number of prognostic signatures for distant metastases, but there has been a limited ability to develop a classifier to predict lymph node status from features of the primary tumour. However, none of the signatures for nodal metastases have been sufficiently robust to have entered clinical practice and do not offer any significant advantages over routine pathological features such as lymphovascular invasion.

A number of studies have also been carried out in other cancers to evaluate primary tumours and lymph node metastases (Hoang et al., 2005, Xi et al., 2005, O'Donnell et al., 2005, Chu et al., 2006). Overall, a number of genes seem to be differentially expressed in association with lymph node metastases including matrix proteins, basement membrane-cell interaction molecules, proteases, cell signalling and oncogenes (Feng et al., 2006, Lee et al., 2003, Hoang et al., 2005, Xi et al., 2005, O'Donnell et al., 2005, Chu et al., 2006).

Patani et al (Patani et al., 2007) reviewed predictors of axillary lymph node status in breast cancer and concluded that no single marker or combination of markers was sufficiently accurate to obviate the need for axillary nodal staging. They also conclude that the aim in future studies must be to reliably identify patients in whom axillary node surgery can be safely omitted.

Cavalli (Cavalli, 2009) also reviewed the molecular markers of breast axillary lymph node metastasis and came to similar conclusions. Clinico-histopathological parameters fail to accurately classify breast cancers according to clinical behaviour. A summary of the most common markers is presented in Table 2 (Cavalli, 2009).

Table 2: Most common markers associated with axillary lymph node breast metastasis.

Method	Type of Assessment	Range of Assessment	Markers	References
IHC	Protein expression	Target proteins	CCND1, CD44, COX2, EGFR, HER2, HPA, LYVE1	(Patani et al., 2007, Park et al., 2007, Cho et al., 2008)
CISH	Gains and amplification/losses and deletions	Target genes/chromosomal regions	CCND1, EGFR, HER2	(Carlsson et al., 2004, Chang et al., 2004, Cho et al., 2008)
RT-PCR	mRNA expression	Target transcripts	CCND1, COX2, CY19, HER2, MAM, MMP2, MMP9, MUC1, NM23, VEGF	(Manzotti et al., 2001, Sakaguchi et al., 2003, Backus et al., 2005, Abdul-Rasool et al., 2006, Revillion et al., 2008)
FISH	Gains and amplification/losses and deletions	Target genes/chromosomal regions	BP1, ER, HER2, PR	(Nathanson et al., 2006, Santinelli et al., 2008, Cavalli et al., 2008)
Expression arrays	mRNA expression	Entire genome	BRF2, CCNE2, CD44, COL9A1, ESM1, FLT1, GSTπ, HER2, MCM6, MMP2, MMP9, MP1, MUC1, MYC, OSF2, RAB6B, RAB11FIP, TOP2A	(Bertucci et al., 2000, van 't Veer et al., 2002, van de Vijver et al., 2002, Huang et al., 2003, Weigelt et al., 2003, Weigelt et al., 2005a, Feng et al., 2006, Hu et al., 2009, Ahr et al., 2001, Ahr et al., 2002, West et al., 2001, Sotiriou et al., 2003, Ramaswamy et al., 2003)
Conventional cytogenetics	Chromosome alterations (numerical and structural)	Entire genome	+1q, del(6)(q23), +8, +9q, del11)(p15), -13, del(17)(p13), -17, -19, -21, -22, del(X)(q25), HSRs	(Emerson et al., 1993, Pandis et al., 1993, Herrington

				et al., 1995, Pandis et al., 1996, Cavalli et al., 1997, Adeyinka et al., 2000, Wuicik et al., 2007, Piao and Malkhosyan, 2002, Teixeira et al., 2002, Zafrani et al., 1992)
CGH	DNA copy number changes (only quantitative)	Entire genome	+1p, -1q, -2q, -4q, -5q, +6p, ++8q, ++9q, +11p, +11q, +12q, -13q, +16, +17p, +17q, -18p, -18q, ++19, ++20q, -21, -22, -Xq	(Kuukasjarvi et al., 1997, Nishizaki et al., 1997, Weber-Mangal et al., 2003, Torres et al., 2007, Friedrich et al., 2008, Cavalli et al., 2003, Santos et al., 2008)
Array CGH	DNA copy number changes (only quantitative)	Entire genome	+1p36, +1q32, +1q41, +6q21, +8q13, +8q24, -11p15, ++11q13, ++17q12, -17p13, -19p13, -19q13, ++20q13, -22q11	(van Beers and Nederlof, 2006, Climent et al., 2007, Chin et al., 2007, Han et al., 2006, Li et al., 2008)
miRNA microarray	Differential expression	Custom miRNA containing 326 human miRNA genes	miR-450a, miR-148a, miR-30b, miR-150, miR-155, miR-99b, miR-125b, miR-205, miR-130b, miR-24, miR-99a	(Baffa et al., 2009)

+: Gain; ++: Amplification; -: Loss; del: Deletion; p: Short arm; q: Long arm

Despite these advances in technology and information about the metastatic process, the accurate prediction of future metastasis remains elusive. Failure to accurately predict which tumours will metastasize shows the inability of histological features to fully reflect the complex molecular biology of breast cancer (Cavalli, 2009). This review concluded that the characterization of genetic alterations present in the first metastatic site in the axilla, the sentinel lymph node (SLN), is a logical step to define the molecular evolution of

primary tumours to a metastatic state, and is probably representative of the initial genetic events that occur in the early metastatic process.

1.10.3. Artificial neural networks and prediction of lymph node metastases in breast cancer

Artificial neural network models (ANN) have been used in a variety of medical applications (Alvager et al., 1994, Astion and Wilding, 1992b, Astion and Wilding, 1992a, Bostwick and Burke, 2001, Burke, 1994, Burke et al., 1997, Cicchetti, 1992, Cross et al., 1995, De Laurentiis and Ravdin, 1994, Egmont-Peterson et al., 1994, Eleuteri et al., 2003, Floyd et al., 1994, Fujita et al., 1992, Hammerston, 1993, Jorgensen et al., 1996, Lundin et al., 1999, Mackiln et al., 1991, Marchevsky et al., 1998, Naguib and Sherbert, 1997, Pedersen et al., 1996, Ravdin and Clark, 1992, Ravdin et al., 1992, Schillen, 1991, Shultz, 1996, Truong et al., 1995, Wilding et al., 1994, Wu et al., 1993, Jerez et al., 2005) and a number of these relate to either survival (Burke et al., 1997, Lundin et al., 1999, Ravdin and Clark, 1992, Ravdin et al., 1992, Jerez et al., 2005) or prediction of lymph node status (Naguib et al., 1997, Grey et al., 2003, Tez et al., 2007, Marchevsky et al., 1999, Naguib et al., 1999) in breast cancer. Kuo et al (S.-J. Kuo, 2008) used a neural network model with three-dimensional power Doppler ultrasound to classify malignant lesions of the breast. The accuracy of the model was 84.6%, sensitivity 90.3%, specificity 79.4%, positive predictive value 80% and the negative predictive value was 90%. Jamarani et al (Jamarani et al., 2005) used a neural network to develop an intelligent system for the identification of microcalcification clusters in digitized mammograms as an aid to radiological diagnosis.

Naguib et al (Naguib et al., 1997) developed an artificial neural network model (ANN) to predict nodal metastases in breast cancer. Using 9 of a possible 12 markers, this model was correctly able to predict axillary involvement in 84% of patients in the test set of 31 patients. Marchevsky et al (Marchevsky et al., 1999) evaluated prognostic factors for prediction of nodal status using an ANN. The model developed correctly classified 89% (40 of 55 unknown cases) with a sensitivity of 80%, specificity of 97.2%, a positive predictive value of 93.8% and a negative predictive value of 87.5%. Bourdes et al (Bourdes et al., 2007) analysed 2535 consecutive breast cancer patients and compared logistic regression with artificial neural network models. Parameters for specific mortality were: sensitivity 80.5% for logistic regression, neural network varLR 86.7% and neural

network varNN 87.5%; specificity 77.9% for logistic regression, neural network varLR 76.0% and neural network varNN 77.3%; false negative rate 19.5% for logistic regression, neural network varLR 13.3% and neural network varNN 12.5%; false positive rate 22.1% for logistic regression, neural network varLR 24.0% and neural network varNN 22.7%; positive predictive value 17.4% for logistic regression, neural network varLR 17.2% and neural network varNN 18.2%. The predictive performance of the neural network for breast cancer-specific mortality and disease-free survival was at least as good as that of logistic regression and significantly better for disease-free survival. The results underline the predictive accuracy of neural network models compared with multivariate linear models and their usefulness as predictive tools. Lisboa et al (Lisboa et al., 2007, Lisboa et al., 2008) used the same database for a time-to-event analysis using neural networks. This study compared neural network modelling to Cox regression for mortality and treatment failure. One of the main advantages of the neural network methodology for modelling time-to-event data is the ability to infer smooth estimates for the hazard without requiring *a priori* assumptions about proportionality (Lisboa et al., 2008). The breast cancer specific mortality study confirmed that successful risk-staging can be performed with both the Cox regression model and the neural network model. The neural network model appeared to be more specific to identify patients at the extremes of high and low risk. For disease-free survival, the neural network model appeared to generalize better than Cox regression.

Lancashire et al (Lancashire et al., 2008) have applied neural network modelling to gene expression profiling data. The amount of data generated by gene expression profiling experiments requires novel data analysis technologies. Lancashire et al (Lancashire et al., 2008) used neural networks to identify an optimal subset of predictive gene transcripts from microarray data. This study applied neural networks to analysis of breast cancers in regards to oestrogen receptor status and axillary lymph node status using a microarray dataset. A logistic regression model gave poor predictive performance with a median accuracy of 78% for ER status and 56% for lymph node status. Using neural network models eight gene transcripts were identified that discriminated ER positive and negative phenotypes with an accuracy of 100% and similarly seven genes were identified to predict lymph node status. Validation was performed using a different dataset and gene transcripts on 89 samples. 88% were correctly classified on ER status with a sensitivity of 90% and a specificity of 80%. 83% of samples were correctly classified on lymph node status with a sensitivity of 87% and a specificity of 80%.

Fuzzy logic algorithms (Seker et al., 2002, Seker et al., 2003) have also been used to predict outcome in breast cancer and Jarman et al (Jarman et al., 2008) have also developed an integrated decision support framework incorporating a number of different mechanisms including the Nottingham Prognostic Index and a rule-extraction method derived from a neural network model.

A gap still exists between the amount of research into breast cancer and the integration of knowledge into diagnostic practice. This proposal aims to develop a practical methodology incorporating new technologies into an approach which directly benefits patients. It incorporates such tested prognostic and predictive biomarkers as oestrogen receptor and pathological features, but intends to enhance the usefulness of these with gene expression profiling targeted at lymphatic spread and nodal metastasis. By this combined approach, it is hoped that this project will advance our understanding of breast cancer progression and begin to enable individualized therapy for patients.

Chapter 2

Materials and Methods.

2.1. Project Outline

The working hypothesis of this project is that the tumour cells present in the primary tumour follow the metastatic cascade by disseminating into lymphatic vessels and then extravasating in lymph nodes to form metastatic deposits. A proportion of the cells derived from the primary tumour are able to form metastatic deposits and if the process is similar to haematogenous dissemination and CTCs then clumps of tumour cells are more likely to form metastases than individual tumour cells. Cells within lymphatic spaces would be expected to have the required molecular and phenotypic characteristics to complete the metastatic pathway to the lymph nodes; to invade into lymphatic channels, survive transit, attach to the lymphatic endothelium, extravasate into the lymph node tissue and form a new tumour deposit in a different environment. Therefore using a combination of different biomarkers, it should be possible to look at characteristics of the primary tumour and possibly the cells in the lymphatic spaces to compare breast carcinomas that metastasis to lymph nodes to those that do not have lymph node metastases.

TMAs will be constructed from a breast cancer database with primary tumour characteristics to evaluate potential biomarkers based on published information. In addition, a sample will be selected with primary tumour, cells from the lymphatic channels and lymph node metastasis to evaluate the feasibility of miRNA sequencing. In the latter case, cellular material will be captured using laser capture microdissection (LCM) from paraffin embedded tissue and subject to miRNA sequencing and discovery performed using an Applied Biosystems SOLiD V4 Next Generation Sequencer.

2.2. Selection of samples

Archival formalin fixed paraffin embedded (FFPE) patient samples have been identified that have clinical characteristics covering a mixture of breast cancer phenotypes including different grades, tumour size and lymph node status. In some cases lymphovascular invasion (LVI) is also identified in the samples.

In 2011, a breast cancer cluster was identified at the ABC studio in Brisbane. Fifteen female patients were diagnosed with breast cancer and an epidemiological analysis of this group of patients indicated that the probability of a causal link was high and it would be

extremely unlikely that this cluster was due to chance. The tumours from these patients were retrieved and examined with a variety of methods including light microscopy, immunohistochemistry, in-situ hybridization and gene expression profiling to ascertain if the tumours were similar to each other and if this cohort would provide information that may be useful for predicting lymph node status. These samples were case matched to samples within the breast cancer database.

2.3. Construction of Database

A web-based database was constructed using Distiller, SlidePath (Leica Microsystems) software. A MindMap was used to determine subgroupings and structure of the database. Three patient cohorts were identified from separate sources and patient records were identified using Systematized Nomenclature of Medicine (SNOMED) coding as a search parameter. Data was extracted into Microsoft Excel and uploaded into Distiller.

3200 patient reports were retrieved from a private pathology database (Data 1), 1834 from a public hospital (Data 2) and 5200 from a separate public pathology hospital (Data 3). Haematoxylin and eosin (H&E) slides from 645 patients were reviewed from Data 1 and pathology information was recorded. This included tumour type, grade and size, presence or absence of lymphatic and vascular permeation, margin type, lymphocytic infiltration, lymph node status and number of involved lymph nodes.

Table 3: Morphological classification of samples used for TMA construction.

	Positive_nodal_status		
Tumour_type_modified			
	N	Y	Total
#na	0	1	1
Cribriform	1	0	1
Ductal	255	186	441
Ductal/Lobular	5	4	9
Lobular	15	14	29
Lobular/Ductal	0	1	1
Mucinous	2	0	2
Papillary	1	1	2
Papillary/Cribriform	1	0	1
Tubulolobular	0	1	1
Total	280	208	488

Table 4: Correlation of tumour grade and lymph node status in samples used for TMA construction.

	Positive_nodal_status		
Grade			
	N	Y	Total
1	111	46	157
2	97	81	178
3	65	80	145
#na	7	1	8
Total	280	208	488

2.4. Construction of Tissue microarrays

A number of tissue microarrays have been constructed using a Beecher ATA-27 automated instrument and a Beecher Galileo semi-automated instrument. TMAs were constructed in duplicate with 0.6 mm cores. Normal control tissue was included in each TMA with cores of liver used for orientation. The archival samples are managed through a SlidePath Distiller and TMAs have been constructed with approximately 1000 patient samples. Scoring sheets have been developed to perform IHC scoring on-line with integration of image analysis for the TMA virtual images.

Tissue microarrays are a relatively new approach to utilize tumour material in a more efficient way (Kononen et al., 1998, Eguiluz et al., 2006). Cores or pieces of tissue are arrayed in a new paraffin block and the presence of multiple tumours in a single block increases the rate of analysis and decreases the costs associated with using multiple antibodies for IHC.
The first multitissue array was described in 1986 by Battifora (Battifora, 1986). Modifications of these techniques have occurred over the last 20 years. The core tissue techniques described by Kononen (Kononen et al., 1998) have been the most widely accepted. The area of interest was identified on a glass microscope slide and the corresponding core was removed from the paraffin block and inserted into the donor block with the two instruments.

The TMA sections have been assessed using immunohistochemistry (IHC) and in situ hybridization (ISH). The TMA slides have been scanned using an automated scanner (Hamamatsu Nanozoomer) and the generated images have the potential to be analysed

using image analysis software. The software can assess nuclear, cytoplasmic and membrane staining as well as morphometry of cells.

Chapter 3

FRANCIS, G. D., JONES, M. A., BEADLE, G. F. & STEIN, S. R. 2009. Bright-field in situ hybridization for HER2 gene amplification in breast cancer using tissue microarrays: correlation between chromogenic (CISH) and automated silver-enhanced (SISH) methods with patient outcome. Diagn Mol Pathol, 18, 88-95.

3.1. Bright-field in-situ hybridization using TMAs

3.1.1. Introduction

Previous studies have shown that TMA methodology is accurate, particularly with breast cancer (Callagy et al., 2003, Camp et al., 2000, Gancberg et al., 2002b, O'Grady et al., 2003, Park et al., 2003, Kay et al., 2004, Simon et al., 2005, Mengel et al., 2003, Zhang et al., 2003, Rui and Lebaron, 2005, Al Kuraya et al., 2004, Ruiz et al., 2005, Watanabe et al., 2005, Warford, 2004, Giltnane and Rimm, 2004, Mobasheri et al., 2004, Torhorst et al., 2001, Aaltonen et al., 2006, Anna et al., 2006). In a recent study, Henriksen et al (Henriksen et al., 2007) assessed the semi-quantitative evaluation of a number of immunohistochemistry assays on 2mm TMAs compared to whole tissue sections. This study showed that ER, PR, AIB1, COX-2, HER2 and IGF-IR on TMA blocks were comparable to whole tissue sections. TMA results for a number of parameters were evaluated and compared to whole tissue sections.

3.1.2. Results

Introduction: *HER2* gene amplification or overexpression occurs in 15-25% of breast cancers and has implications for treatment and prognosis. The most commonly used methods for *HER2* testing are fluorescence in-situ hybridization (FISH) and immunohistochemistry (IHC). FISH is considered to be the reference standard and more accurately predicts response to trastuzumab, but is technically demanding, expensive and requires specialized equipment. In-situ hybridization is required to be eligible for adjuvant treatment with trastuzumab in Australia. Bright Field in-situ hybridization (BRISH) is an alternative to FISH and uses a combination of in-situ methodology and a peroxidase-mediated chromogenic substrate such as diaminobenzidine (CISH™) or multimer technology coupled with enzyme metallography (SISH) to create a marker visible under bright field microscopy. CISH™ was introduced into diagnostic testing in Australia in October 2006. SISH methodology is a more recent introduction into the testing repertoire. An evaluation of CISH™ and SISH performance to assess patient outcome was performed using tissue microarrays.

Materials and Methods: Tissue microarrays were constructed in duplicate using material from 593 patients with invasive breast carcinoma and assessed using CISH™ and SISH.

Gene amplification was assessed using the American Society of Clinical Oncology/College of American Pathologists guideline and Australian *HER2* Advisory Board criteria (Single probe: Diploid, 1-2.5 copies/nucleus; polysomy >2.5-4 copies/nucleus; Equivocal, >4-6 copies/nucleus; low level amplification, >6 -10 copies/nucleus and high level amplification >10 copies/nucleus; Dual probe *HER2*/CHR17 ratio: Non-amplified <1.8, Equivocal 1.8-2.2, Amplified >2.2).

Results:

Results were informative for 337 tissue cores comprising 230 patient samples. Concordance rates were 96% for *HER2* single probe CISH and SISH and 95.5% for single probe CISH™ and dual probe *HER2*/CHR17 SISH. Both bright field methods correlated with IHC results and with breast cancer specific survival.

Conclusion:

HER2 SISH testing combines the advantages of automation and bright field microscopy to facilitate workflow within the laboratory, improves turnaround time, and correlates with patient outcome.

Introduction

ERBB2 (HER2) amplification was originally described as occurring in 20-25% of breast carcinomas in 1989(Slamon et al., 1989).

With the development of targeted therapy in both metastatic and adjuvant settings the assessment of *HER2* status is critical in the treatment of breast cancer patients. Trastuzumab, a humanised mouse monoclonal antibody was developed to target tumours with overexpression of *HER2*. Other targeted therapy such as the tyrosine kinase inhibitor, lapatinib (Rusnak et al., 2001, Chang, 2007), is in clinical trials.

HER2 overexpression has been linked to a poor prognosis in breast carcinoma patients (Slamon et al., 1987, Gusterson et al., 1992, Ferretti et al., 2007, Peiro et al., 2007a, Ryden et al., 2007). Most studies have indicated overexpression is associated with a poor response to hormonal therapy and that these patients receive increased benefit from anthracycline-based chemotherapy (Thor et al., 1998, Paik and Park, 2001, Paik et al., 1998, Masood and Bui, 2002, Dhesy-Thind et al., 2007).

HER2 testing can be performed by a number of methods, either evaluating protein expression or gene copy number. The two most common methods are immunohistochemistry (IHC) and fluorescence in-situ hybridization (FISH) (van de Vijver, 2002). A number of variables including fixation time, antibody selection, retrieval time and laboratory technique (Gancberg et al., 2002a, Wolff et al., 2007) will have an impact on the results obtained by these techniques. Inter-observer interpretative variability also occurs (Pathologists, 2002).

The majority of pathology laboratories are able to perform IHC and this methodology is relatively inexpensive and easily integrated into routine diagnostic work. Turnaround time is rapid and the method can be automated. The interpretation of the staining is subjective and at best semiquantitative.

The gold (reference) standard for *HER2* evaluation is considered to be FISH, however it is technically demanding, expensive, does not produce a permanent archival slide, and requires specialized equipment (Bartlett et al., 2001, Diaz, 2001).

A number of studies have noted discordant results between FISH and IHC results (Bartlett et al., 2001, Konecny and Slamon, 2002, Perez et al., 2002). To further confuse the issue, the initial clinical correlations were performed with an IHC antibody that is no longer available (Perez et al., 2002). Logically, it would seem that protein expression would more accurately reflect the response to targeted therapy, but a number of studies have indicated

that gene amplification is a more accurate predictor (Elkin and Schnitt, 2004, Mass et al., 2005, Riou et al., 2001).

Alternative methodologies have been developed because of these constraints. These include enzyme-linked immunosorbent assay for the extracellular domain in serum, Bright field in-situ hybridization (BRISH) such as chromogenic in-situ hybridization (CISH) and real-time polymerase chain reaction (van de Vijver, 2002, Zhao et al., 2002, Dandachi et al., 2002).

BRISH uses a combination of in-situ hybridization with a detection system using a staple chromogen similar to IHC. The slides are visible under bright field microscopy and show correlation with tumour morphology. Extensive correlation studies have been performed with CISH and both IHC and FISH (Dandachi et al., 2002, Zhao et al., 2002, Hauser-Kronberger and Dandachi, 2004, Peiro et al., 2007b). CISH shows very good correlation with FISH (Arnould et al., 2003, Gupta et al., 2003, Park et al., 2003, Bilous et al., 2006). In addition to the CISH methodology a number of other BRISH techniques are available. These include silver autometallography (Sinczak-Kuta et al., 2007), dual colour ISH (Laakso et al., 2006, Shipley, 2006) and automated silver enhanced in-situ hybridization (SISH)(Dietel et al., 2007). A recent study has been performed comparing the performance of FISH with SISH in 99 patients (Dietel et al., 2007) and an evaluation of ISH methodology by Sinczak-Kuta(Sinczak-Kuta et al., 2007) indicated that they are equivalent tools for evaluating *HER2* gene amplification. A small number of studies have evaluated the prognostic significance of *HER2* gene amplification by CISH(Dandachi et al., 2002, Joensuu et al., 2003). Two recent studies evaluated the prognostic implications of *HER2* status in lymph node negative breast cancer using CISH(Peiro et al., 2007a, Peiro et al., 2007b) and a third study used CISH to assess the prognostic significance of *HER2* gene amplification in metastatic breast cancer. The current study was performed to evaluate the clinical performance of SISH and CISH, both bright field methods with patient outcome in primary breast cancer.

Methods

Breast carcinoma patients were retrieved from a database between 1990 and 2002. Tissue microarrays (TMAs) were constructed with material from 589 patients. Duplicate 0.6mm x 0.6mm cores were removed from the donor blocks and placed into a total of 8 recipient blocks. All slides were reviewed by a single pathologist to confirm the

morphological features of the tumours. The study used paraffin embedded blocks of tumour that had undergone fixation in 10% buffered formalin and routine processing. The fixation protocols and processing times varied.

HER2 Immunohistochemistry

IHC was performed on the TMAs using the HercepTest™ (Dako, CA).

Tissue sections were dewaxed in xylene, rehydrated in graded alcohols followed by antigen retrieval in Citrate buffer pH6.0 as specified by the manufacturer. Counter staining was performed with haemotoxylin.

IHC staining was evaluated semiquantitatively using the HercepTest™ system for HER-2. In Australia, a score of 3+ is regarded as positive, 2+ is equivocal and and 1+ and 0 are negative.

1.11. *HER2* Chromogenic in-situ hybridization

HER2 CISH was performed using Invitrogen (Carlsbad, CA), CISH pre-treatment kit; CISH *HER2* probe and Immunodetection Kit. Sections were de-paraffinised and HIER was performed using Pre-treatment 1 solution from the Invitrogen pre-treatment kit in Biocare's Decloaking Chamber followed by incubation in Pre-treatment 2 enzyme. Sections were de-hydrated and HER-2 probe added to the section and sealed using a glass coverslip and liquid cement. A thermal cycler was used for probe hybridization followed by overnight incubation at 37 deg C. On day 2 Immunodetection was performed using the Invitrogen CISH Immunodetection kit after a stringency wash according to the manufacturer's instructions. Gene amplification was assessed using the American Society of Clinical Oncology/College of American Pathologists guideline and Australian *HER2* Advisory Board criteria (Diploid, 1-2.5 copies/nucleus; polysomy >2.5-4 copies/nucleus; Equivocal, >4-6 copies/nucleus; low level amplification, >6 -10 copies/nucleus and high level amplification >10 copies/nucleus).

HER2 Silver enhanced in-situ hybridization

HER2 SISH was performed on an automated instrument, Ventana Benchmark® (Ventana Medical Systems, Tucson, AZ) as per the manufacturer's protocols for the INFORM™ *HER2* DNA probe and Chromosome 17 probes. Testing for the *HER2* gene and chromosome 17 was performed on sequential sections. Both probes are labeled with dinitrophenol. Denaturation occurs on the instrument with enzyme digestion in Protease 3

for 8 minutes. The detection system uses a multimer labeled with goat anti-rabbit antibody horse radish peroxidase as the linking step. Visualization occurs with the sequential addition of silver acetate as the source of ionic silver, hydroquinone and hydrogen peroxide to give a black metallic silver precipitate at the probe site. Counterstaining is performed with Hamatoxilin II on the instrument. The time taken for the complete run is 6.5 hours. Both *HER2* and Chromosome 17 detection can be performed on the same slide run. Gene amplification was assessed using the American Society of Clinical Oncology/College of American Pathologists guideline and Australian *HER2* Advisory Board criteria for single *HER2* probe testing (Diploid, 1-2.5 copies/nucleus; polysomy >2.5-4 copies/nucleus; Equivocal, >4-6 copies/nucleus; low level amplification, >6 -10 copies/nucleus and high level amplification >10 copies/nucleus) and for dual *HER2*/CHR17 probe testing (Non amplified ratio <1.8; Equivocal ratio, 1.8-2.2; Gene amplification, > 2.2).

Scores were counted by a single pathologist (GF) with blinding to patient outcome and IHC scores.
Statistical analysis and Kaplan-Meier plots were performed using SPSS 15.0 for Windows (SPSS Inc, Chicago IL).

Results:
Results were available on 337 tissue cores and 230 patient samples for both methods. Two of the 339 cores containing tumour (0.6%) were non-informative with SISH testing. Twenty-four of the 428 cores containing tumour (5.6%) were non-informative with CISH testing. The difference in core numbers for CISH and SISH was due to tissue loss due to block sectioning. Approximately 20 sections were utilized between the performance of the CISH and subsequent SISH testing.
Tumour features of the cases are listed in Table 1. The median time of patient follow up is 139.3 months.
IHC correlation was similar for single probe CISH and single probe SISH (Tables 2 & 3). A single discrepant result was identified with the dual probe *HER2*/Chr17 ratio with one case showing 3+ IHC staining but a normal *HER2*/Chr17 ratio (Table 4). Images of representative TMA cores are illustrated in Figures 1 & 2 for IHC, CISH and SISH. Correlation between individual cores and patient samples was high for the single probe CISH and SISH with Pearson's R 0.945, R-Square 0.89 and Spearman correlation 0.747 (Figure 3). Using the five subcategories for classification of *HER2* status with both single

probes, there was agreement in 88% of patient specimens (Table 5). Compacting the subcategories into 3 clinically significant categories; diploid/polysomy, equivocal and amplified, there was agreement in 96% of cases (Table 6). Discrepancies occurred in the equivocal categories in both methods. When single probe CISH was compared to the SISH *HER2*/chr 17 ratio the concordance for the clinically significant categories was 95.6% (Table 7). Again, discrepancies occurred in the equivocal groups. In the single case indicated above, single probe results showed low *HER2* gene amplification with both CISH and SISH, but the *HER2*/Chr17 ratio was within the normal range indicating high level polysomy.

A Bland-Altman analysis (Figure 4) shows that there is no evidence of bias between the 2 methods (mean difference = -0.002, 95% confidence interval: -0.19 to 0.19) and in most cases the difference between the 2 methods will be a maximum count of +/- 2.9.

Patient outcome was evaluated using Kaplan-Meier plots for breast cancer specific survival. The plots show virtually identical survival statistics (Figure 5).

Discussion:

HER2 testing has developed over a number of years, and whilst the role of *HER2* as a predictive biomarker is now relatively well defined, the role as a prognostic marker is based on retrospective studies with inconsistent results(Ferretti et al., 2007, Todorovic-Rakovic et al., 2007).

In Australia from the 1st October 2006, ISH testing is required for patients to be eligible for adjuvant trastuzumab in conjunction with chemotherapy. The intent is that all new cases are tested by ISH for *HER2*. The rationale for this decision is based on the results for Quality Assurance testing for *HER2* IHC in Australia(Francis et al., 2007b). IHC may still be performed as an additional test. Some laboratories in Australia currently test all new primary breast cancer cases with ISH. However, as an interim measure patients can still be triaged using IHC and then have ISH testing performed on all 2+ and 3+ cases to confirm *HER2* gene amplification. Cases with 1+ IHC staining can also be tested by ISH if there is a suspicion that the IHC is incorrect. This testing algorithm was developed as there were insufficient laboratories that were able to introduce ISH testing on all new primary breast cancer cases within the timeframe required for implementation. There are currently 19 laboratories in Australia that perform CISH testing and one laboratory that performs FISH testing for *HER2*.

The correlation studies indicate that the two bright field methods are similar with a concordance rate above 95% as specified in the ASCO/CAP *HER2* testing guidelines (Wolff et al., 2007) and this correlation is similar to the previous ISH evaluation(Sinczak-Kuta et al., 2007). The Bland-Altman plot shows no evidence of bias in the two methods. However, whilst correlation between technical methods is useful, ultimately it is the ability of the test to define patient groups that is the most critical. Kaplan-Meier plots for breast cancer specific survival curves were generated using both single probe assays and the dual probe SISH assay in addition to the results obtained for *HER2* IHC testing. The two bright field ISH methods generate curves showing virtually identical results. Whilst the groups of patients are not corrected for hormone receptor status and treatment, the curves show that *HER2* gene amplified tumours have a worse outcome compared to *HER2* negative tumours. This indicates the clinical usefulness of these assays in evaluating *HER2* status in primary breast cancer. Kaplan-Meier survival curves were also generated using the *HER2* IHC results. There were three categories for staining. The curves are similar to the ISH curves with the IHC negative (0 & 1+ staining) cases overlying the curve for non-amplified cases. The patients with equivocal (2+ staining) and positive (3+ staining) tumours showed slightly divergent outcomes. No patient received trastuzumab within these groups. It could be argued from these results that *HER2* IHC is an acceptable method of predicting *HER2* status as defined by patient outcome. However, in diagnostic laboratories, the results are more variable. An audit was performed for laboratories performing HER IHC testing in Australasia (Francis et al., 2007b). Whilst the average results for *HER2* 3+ staining fell within acceptable levels (17.1%), there was marked variation in individual laboratory's results from 5% to more than 60% reported as positive. With bright field ISH testing there are a small number of equivocal results with a *HER2* single probe method require testing using a dual probe method, either dual probe bright field ISH or FISH. Equivocal cases with single probe CISH therefore would be tested with a dual probe system and the results from this methodology would be issued as the final report. Equivocal results on a dual probe system require full evaluation of both the mean *HER2* cell count and the IHC results for interpretation. There will however be a small number of patients where the tumour response to trastuzumab will be uncertain and testing for protein and gene expression by all available methods will fall into the equivocal ranges.

From a workflow perspective, SISH offers the advantage of automation with a 6 hour protocol. CISH requires an overnight incubation and is manual with increased labour

requirements. The automation is more readily integrated into the laboratory's processes, but does require a Ventana instrument to perform the analysis.

Conclusion:

BRISH testing for *HER2* gene amplification in breast cancer has recently been developed to enable diagnostic laboratories to perform this assessment in the routine work processes. SISH is a recently developed method that offers the advantages of bright field ISH coupled with automation for *HER2* gene amplification testing. It has a high concordance with CISH and FISH (Dietel et al., 2007) and fulfills the requirements for diagnostic testing. *HER2* SISH correlation with *HER2* IHC is similar to that of CISH in this study. Both of these bright field methods show a correlation with breast cancer specific survival in this patient cohort confirming the clinical utility of this methodology.

A Quality Assurance Program has been developed for bright field *HER2* ISH testing through the Royal College of Pathologists of Australasia Quality Assurance Program. This program incorporates both *HER2* CISH and *HER2* SISH methodology. This will enable the introduction of this testing methodology into diagnostic laboratories.

References:

1. Slamon, D., W. Godolphin, L. Jones, et al., *Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer.* Science, 1989. **244**: p. 707-712.
2. Rusnak, D.W., K. Affleck, S.G. Cockerill, et al., *The characterization of novel, dual ErbB-2/EGFR, tyrosine kinase inhibitors: potential therapy for cancer.* Cancer Res, 2001. **61**(19): p. 7196-203.
3. Chang, J.C., *HER2 Inhibition: From Discovery to Clinical Practice.* Clin Cancer Res, 2007. **13**(1): p. 1-3.
4. Slamon, D., G. Clark, S. Wong, et al., *Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene.* Science, 1987. **235**: p. 177-182.
5. Gusterson, B.A., R.D. Gelber, and A. Goldhirsch, *Prognostic importance of c-erbB-2 expression in breast cancer.* J Clin Oncol, 1992. **10**(1049-1056).
6. Ferretti, G., A. Felici, P. Papaldo, et al., *HER2/neu role in breast cancer: from a prognostic foe to a predictive friend.* Curr Opin Obstet Gynecol, 2007. **19**(1): p. 56-62.

7. Peiro, G., E. Adrover, F.I. Aranda, et al., *Prognostic implications of HER-2 status in steroid receptor-positive, lymph node-negative breast carcinoma.* Am J Clin Pathol, 2007. **127**(5): p. 780-6.
8. Ryden, L., G. Landberg, O. Stal, et al., *HER2 status in hormone receptor positive premenopausal primary breast cancer adds prognostic, but not tamoxifen treatment predictive, information.* Breast Cancer Res Treat, 2007.
9. Thor, A., D. Berry, and D. Budman, *ErbB-2, p53, and eficacy of adjuvant therapy in lymph-node positive breast cancer.* J Natl Cancer Inst, 1998. **90**(18): p. 1346-1360.
10. Paik, S. and C. Park, *HER-2 and choice of adjuvant chemotherapy in breast cancer.* Seminars in Oncology, 2001. **28**(4): p. 332-335.
11. Paik, S., J. Bryant, C. Park, et al., *erb-2 and response to Doxorubicin in patients with axillary lymph node-positive, hormone receptor-negative breas cancer.* Journal of the National Cancer Institute, 1998. **90**(18): p. 1361-1370.
12. Masood, S. and M.M. Bui, *Prognostic and predictive value of HER2/neu oncogene in breast cancer.* Microscopy Research and Technique, 2002. **59**: p. 102-108.
13. Dhesy-Thind, B., K.I. Pritchard, H. Messersmith, et al., *HER2/neu in systemic therapy for women with breast cancer: a systematic review.* Breast Cancer Res Treat, 2007.
14. van de Vijver, M., *Emerging technologies for HER2 testing.* Oncology, 2002. **63(suppl 1)**: p. 33-38.
15. Gancberg, D., T. Jarvinen, A. Di Leo, et al., *Evaluation of HER-2/NEU protein expression in breast cancer by immunohistochemistry: an interlaboratory study assessing the reproducibility of HER-2/NEU testing.* Breast Cancer Research and Treatment, 2002. **74**: p. 113-120.
16. Wolff, A.C., M.E. Hammond, J.N. Schwartz, et al., *American Society of Clinical Oncology/College of American Pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer.* J Clin Oncol, 2007. **25**(1): p. 118-45.
17. Pathologists, C.o.A., *Clinical laboratory assays for HER-2/neu amplification and overexpression: Quality assurance, standardization, and proficiency testing.* Arch Pathol Lab Med, 2002. **126**: p. 803-808.
18. Bartlett, J., J. Going, E. Mallon, et al., *Evaluating HER2 amplification and overexpression in breast cancer.* Journal of Pathology, 2001. **195**: p. 422-428.

19. Diaz, N.M., *Laboratory Testing for HER2/neu in Breast Carcinoma: An Evolving Strategy to Predict Response to Targeted Therapy.* Cancer Control, 2001. **8**(5): p. pp 415 - 418.
20. Konecny, G. and D. Slamon, *HER2 testing and correlation with efficacy of Trastuzumab therapy.* Oncology, 2002. **16**(11).
21. Perez, E.A., P. Roche, R. Jenkins, et al., *HER2 testing in patients with breast cancer:Poor correlation between weak positivity by immunohistochemistry and gene amplification by fluorescence in situ hybridization.* Mayo Clin Proc, 2002. **77**: p. 148-154.
22. Elkin, E.B. and S.J. Schnitt, *In Reply.* J Clin Oncol, 2004. **22**(20): p. 4232-a-4233.
23. Mass, R.D., M.F. Press, S. Anderson, et al., *Evaluation of clinical outcomes according to HER2 detection by fluorescence in situ hybridization in women with metastatic breast cancer treated with trastuzumab.* Clin Breast Cancer, 2005. **6**(3): p. 240-6.
24. Riou, G., M.-C. Mathieu, M. Barrois, et al., *c-erb-2 (HER2/neu) gene amplification is a better indicator of poor prognosis than protein over-expression in operable breast-cancer patients.* Int J Cancer (Pred Oncol), 2001. **95**: p. 266-270.
25. Zhao, J., R. Wu, A. Au, et al., *Determination of HER2 gene amplification by chromogenic in situ hybridization (CISH) in archival breast carcinoma.* Mod Pathol, 2002. **15**(6): p. 657-665.
26. Dandachi, N., O. Dietze, and C. Hauser-Kronberger, *Chromogenic in situ hybridization: A novel approach to a practical and sensitive method for the detection of HER2 oncogene in archival human breast carcinoma.* Laboratory Investigation, 2002. **82**(8): p. 1007-1014.
27. Hauser-Kronberger, C. and N. Dandachi, *Comparison of chromogenic in situ hybridization with other methodologies for HER2 status assessment in breast cancer.* J Mol Histol, 2004. **35**(6): p. 647-53.
28. Peiro, G., F.I. Aranda, E. Adrover, et al., *Analysis of HER2 by chromogenic in situ hybridization and immunohistochemistry in lymph node-negative breast carcinoma: Prognostic relevance.* Hum Pathol, 2007. **38**(1): p. 26-34.
29. Arnould, L., Y. Denoux, G. MacGrogan, et al., *Agreement between chromogenic in situ hybridisation (CISH) and FISH in the determination of HER2 status in breast cancer.* British J of Cancer, 2003. **88**: p. 1587-1591.

30. Gupta, D., L.P. Middleton, M.J. Whitaker, et al., *Comparison of fluorescence and chromogenic in situ hybridization for detection of HER-2/neu oncogene in breast cancer.* Am J Clin Pathol, 2003. **119**(3): p. 381-7.
31. Park, K., J. Kim, S. Lim, et al., *Comparing fluorescence in situ hybridization and chromogenic in situ hybridization methods to determine the HER2/neu status in primary breast carcinoma using tissue microarray.* Mod Pathol, 2003. **16**(9): p. 937-43.
32. Bilous, M., A. Morey, J. Armes, et al., *Chromogenic in situ hybridisation testing for HER2 gene amplification in breast cancer produces highly reproducible results concordant with fluorescence in situ hybridisation and immunohistochemistry.* Pathology, 2006. **38**(2): p. 120-4.
33. Sinczak-Kuta, A., R. Tomaszewska, L. Rudnicka-Sosin, et al., *Evaluation of HER2/neu gene amplification in patients with invasive breast carcinoma. Comparison of in situ hybridization methods.* Pol J Pathol, 2007. **58**(1): p. 41-50.
34. Laakso, M., M. Tanner, and J. Isola, *Dual-colour chromogenic in situ hybridization for testing of HER-2 oncogene amplification in archival breast tumours.* J Pathol, 2006. **210**(1): p. 3-9.
35. Shipley, J., *Putting the colours into chromogenic in situ hybridization (CISH).* J Pathol, 2006. **210**(1): p. 1-2.
36. Dietel, M., I.O. Ellis, H. Hofler, et al., *Comparison of automated silver enhanced in situ hybridisation (SISH) and fluorescence ISH (FISH) for the validation of HER2 gene status in breast carcinoma according to the guidelines of the American Society of Clinical Oncology and the College of American Pathologists.* Virchows Arch, 2007. **451**(1): p. 19-25.
37. Joensuu, H., J. Isola, M. Lundin, et al., *Amplification of erbB2 and erbB2 expression are superior to estrogen receptor status as risk factors for distant recurrence in pT1N0M0 breast cancer: a nationwide population-based study.* Clin Cancer Res, 2003. **9**(3): p. 923-30.
38. Todorovic-Rakovic, N., D. Jovanovic, Z. Neskovic-Konstantinovic, et al., *Prognostic value of HER2 gene amplification detected by chromogenic in situ hybridization (CISH) in metastatic breast cancer.* Exp Mol Pathol, 2007. **82**(3): p. 262-8.
39. Francis, G.D., M. Dimech, L. Giles, et al., *Frequency and reliability of oestrogen receptor, progesterone receptor and HER2 in breast carcinoma determined by*

immunohistochemistry in Australasia: results of the RCPA Quality Assurance Program. J Clin Pathol, 2007. **60**(11): p. 1277-1283.

Clinicopathological characteristics of Patients with Breast Cancer		
Characteristic	n	%
Age at Diagnosis		
50 years or less	74	32.2
> 50 years	155	67.4
Unknown	1	0.4
Menopausal status		
Premenopausal	27	11.7
Pregnant	1	0.4
Perimenopausal	8	3.5
Postmenopausal	117	50.1
Unknown	77	33.5
Tumour type		
Ductal NOS	200	87
Ductal/Lobular	6	2.6
Lobular	14	6.1
Cribriform	1	0.4
Mucinous	2	0.9
Papillary	1	0.4
Papillary/Cribriform	1	0.4
Tubulolobular	1	0.4
Unknown	4	1.7
Histological Grade		
GI	78	33.9
GII	91	39.6
GIII	51	22.2
Unknown	10	4.3
Tumour Size		
<10mm	2	0.9
10-20mm	121	52.6
>20-50mm	80	34.8
>50mm	8	3.5
Unknown	19	8.3
Hormone receptor status		

ERp+PRp+	147	63.9
ERp+PRp-	24	10.4
Erp-PRp+	15	6.52
Erp-PRp-	44	19.1
Lymph node status		
Positive	78	33.9
Negative	97	42.2
Unknown	55	23.9
Patient outcome		
Death from disease	59	25.7
Death from other causes	26	11.3
Alive	143	62.2
Death from unknown cause	2	0.9

Table 1: Clinicopathological characteristics of patient cohort.

IHC *HER2* status	CISH *HER2* status					
	Diploid	Polysomy	Equivocal	Low amplification	High amplification	**Total**
0	106	6	2	0	0	**114**
1	60	10	3	0	1	**74**
2	3	3	1	5	5	**17**
3	0	0	0	4	19	**23**
Total	**169**	**19**	**6**	**9**	**25**	**228**

Table 2: Comparative data for *HER2* IHC (HercepTest™) and CISH status.

IHC *HER2* status	SISH *HER2* status					
	Diploid	Polysomy	Equivocal	Low amplification	High amplification	**Total**
0	105	7	0	1	0	**113**
1	60	13	0	0	1	**74**
2	4	0	3	3	7	**17**
3	0	0	0	3	20	**23**
Total	**169**	**20**	**3**	**7**	**28**	**227**

Table 3: Comparative data for *HER2* IHC (HercepTest™) and SISH *HER2* gene copy status.

IHC *HER2* status	SISH *HER2*/CHR17 ratio status				
	Diploid	Equivocal	Low amplification	High amplification	**Total**
0	111	1	0	0	**112**
1	70	1	0	1	**72**
2	5	1	3	7	**16**
3	1	0	1	21	**23**

Total	**187**	**3**	**4**	**29**	**223**

Table 4: Comparative data for *HER2* IHC (HercepTest™) and SISH *HER2*/CHR17 ratio status.

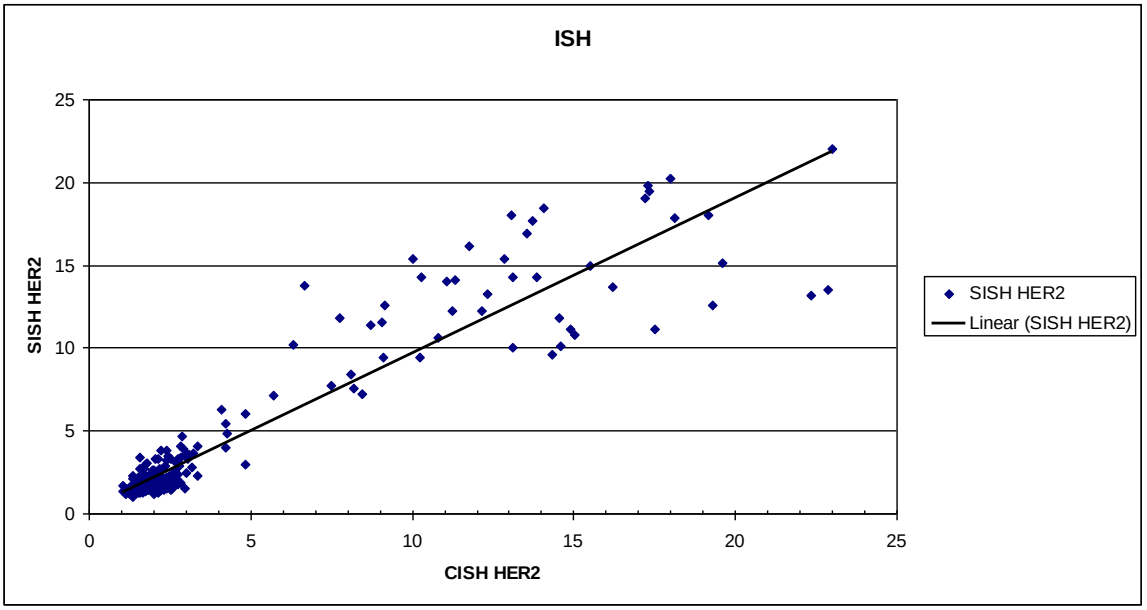


Figure 3: Scatterplot of results for 337 individual TMA cores for *HER2* gene copy number comparing CISH and SISH.

CISH *HER2* status	SISH *HER2* status					
	Diploid	Polysomy	Equivocal	Low amplification	High amplification	**Total**
Diploid	163	8	0	0	0	**170**
Polsomy	6	11	2	0	0	**19**
Equivocal	3	1	0	2	0	**6**
Low amplification	0	0	1	4	4	**9**
High amplification	0	0	0	1	24	**25**
Total	**172**	**20**	**3**	**7**	**28**	**230**

Table 5: Comparative data for *HER2* gene amplification status based on average mean cell count for patient samples on TMA cores.

CISH *HER2* status	SISH *HER2* status			
	Diploid/Polysomy	Equivocal	Amplification	**Total**
Diploid/Polysomy	188	2	0	**189**
Equivocal	4	0	2	**6**
Low amplification	0	1	33	**34**
Total	**192**	**3**	**35**	**230**

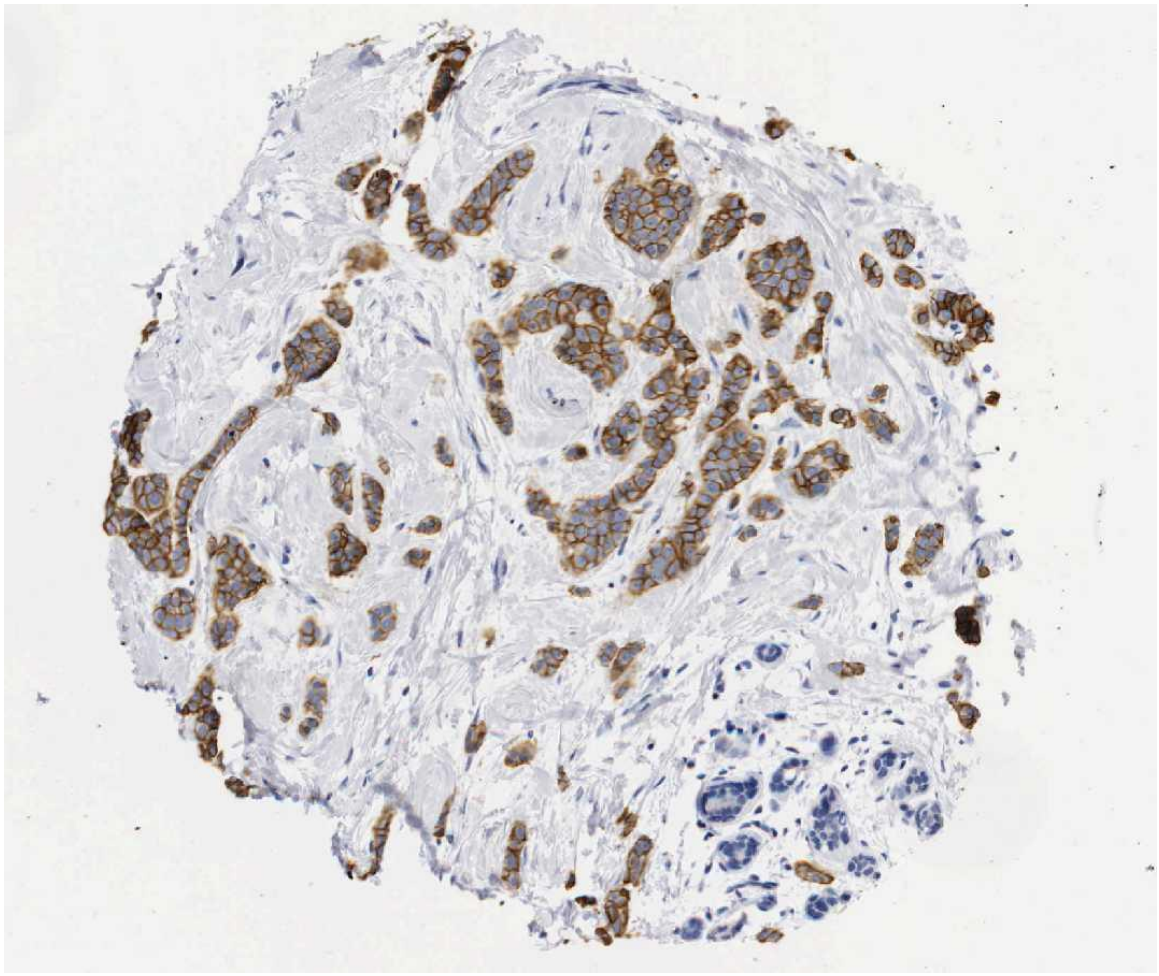
Table 6: Comparative data for *HER2* gene amplification status based on average mean cell count for patient samples on TMA cores with clinical groups.

CISH *HER2* status	SISH *HER2*/CHR17 ratio status				
	Diploid	Equivocal	Low amplification	High amplification	**Total**
Diploid	165	2	0	0	**167**
Polsomy	18	1	0	0	**19**
Equivocal	5	0	1	0	**6**
Low amplification	1	0	2	5	**8**
High amplification	0	0	1	24	**25**
Total	**189**	**3**	**4**	**29**	**225**

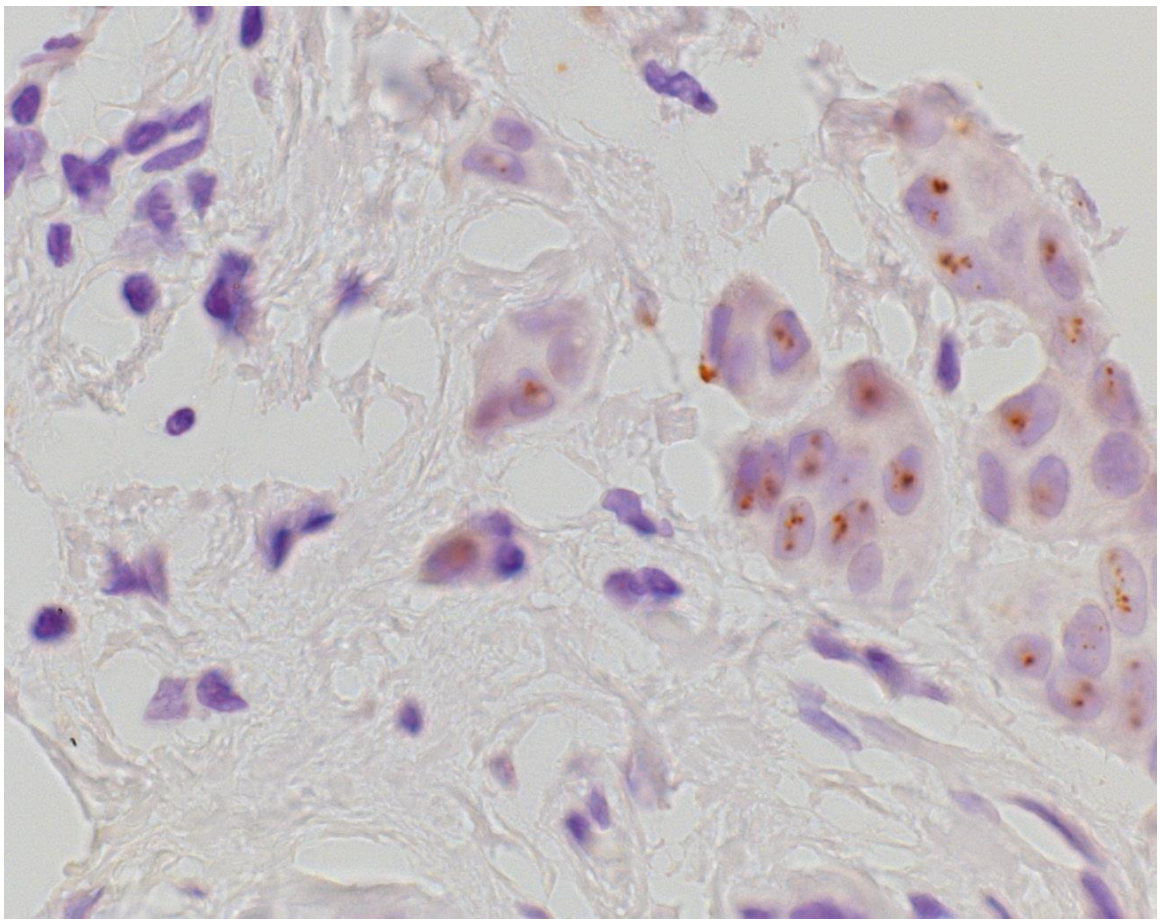
Table 7: Comparative data for *HER2* gene amplification status based on average CISH mean cell count for patient samples on TMA cores and SISH *HER2*/CHR17 ratio.

Figure 1: Representative TMA *HER2* staining with IHC, CISH and SISH.

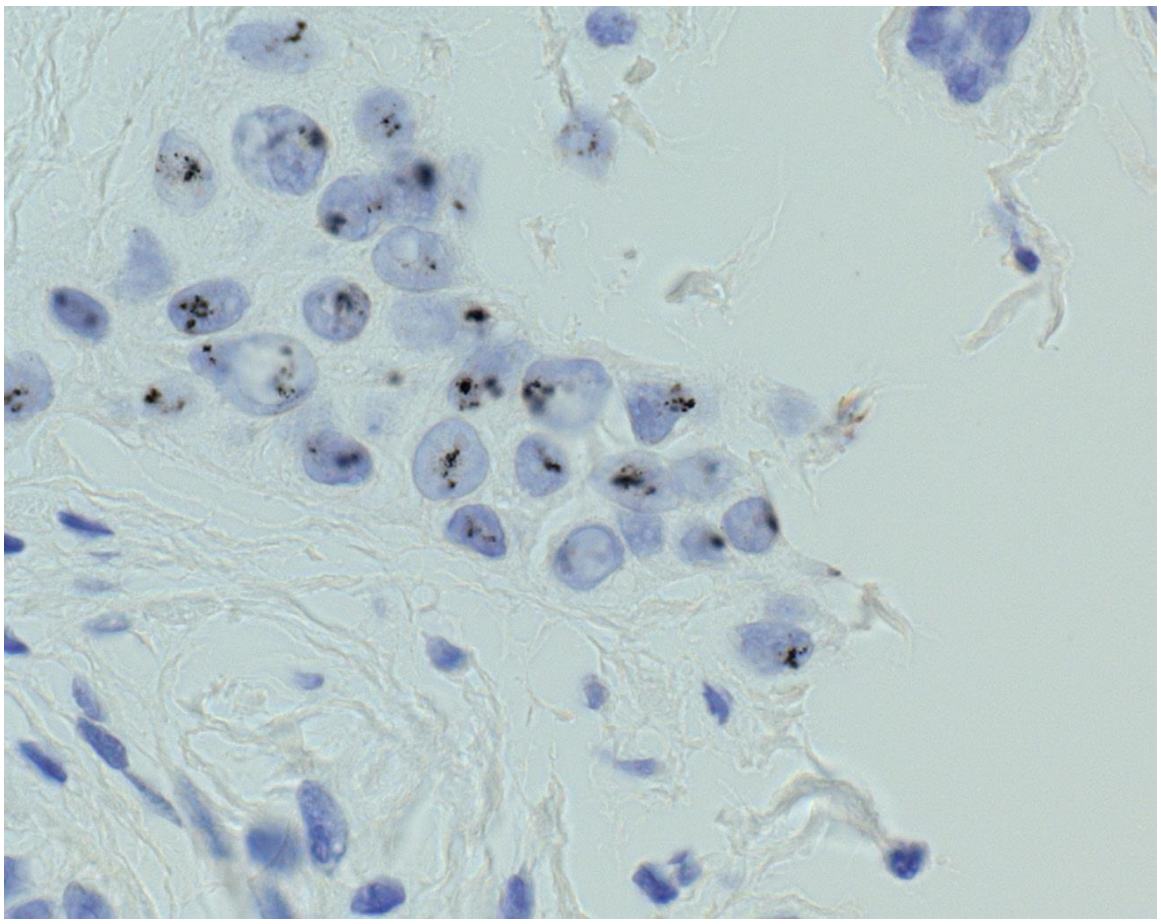
A) IHC 2+ membrane staining (Aperio ScanScope® T2 x20)
B) CISH, Gene amplification (Zeiss AxioImager x630)
C) SISH HER2 gene amplification (Zeiss AxioImager x630)
D) SISH Chr 17 (Zeiss AxioImager x630)



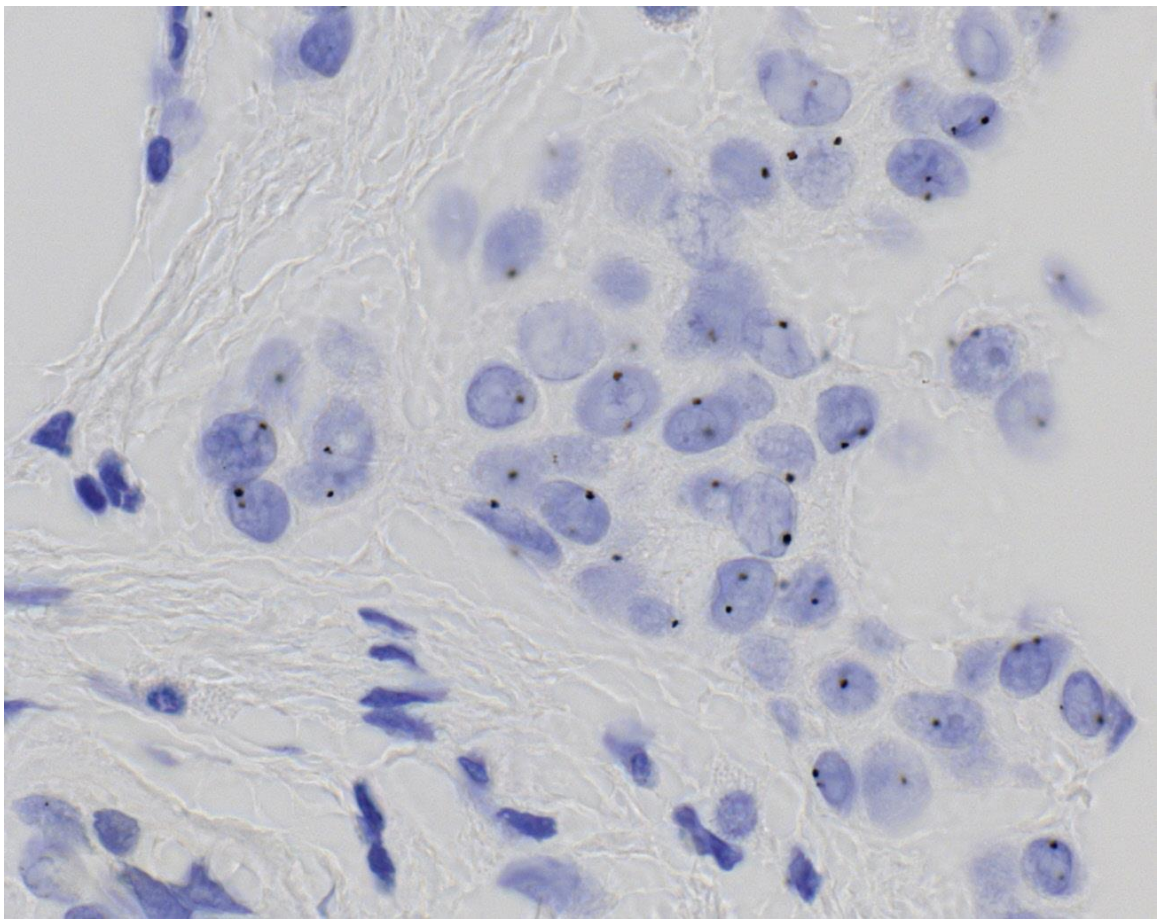
A.



B.



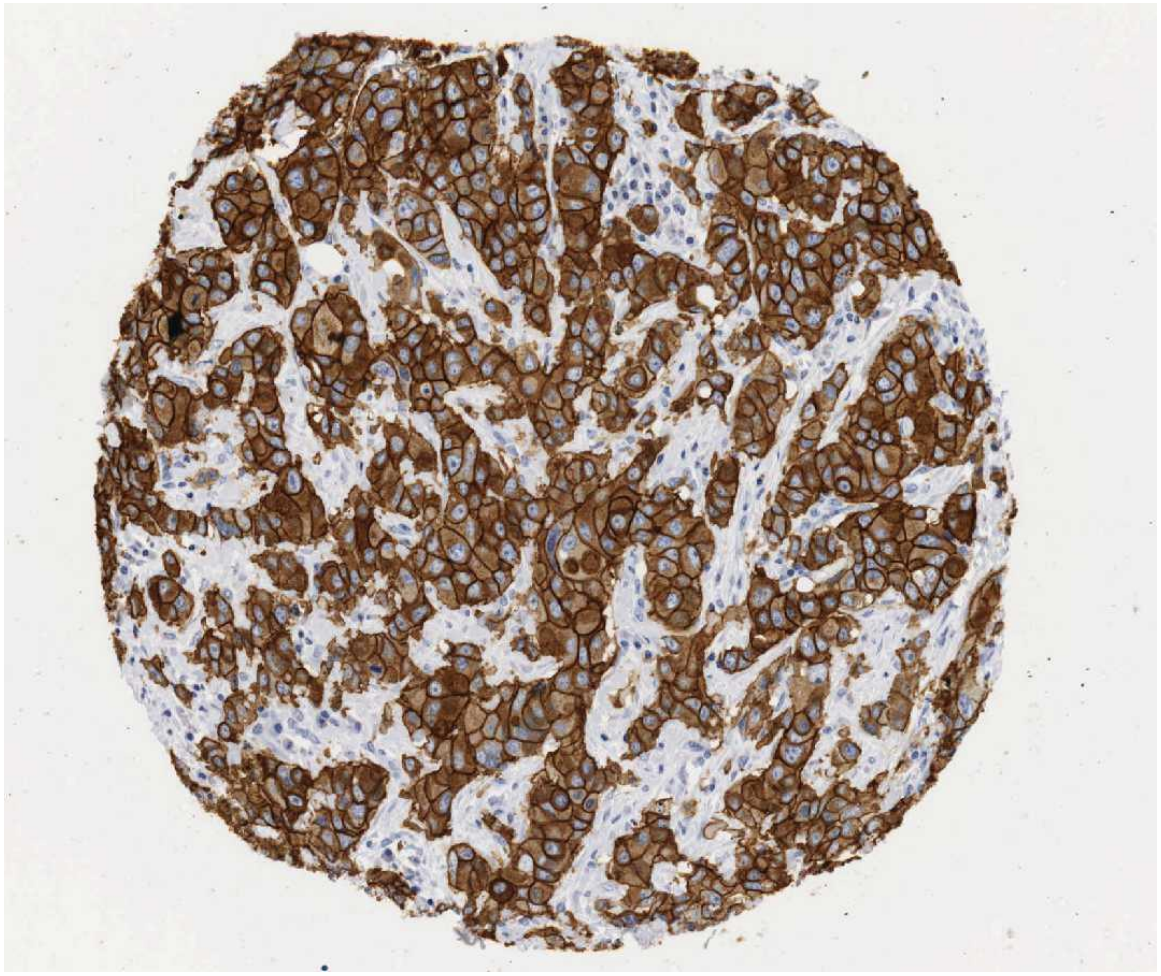
C.



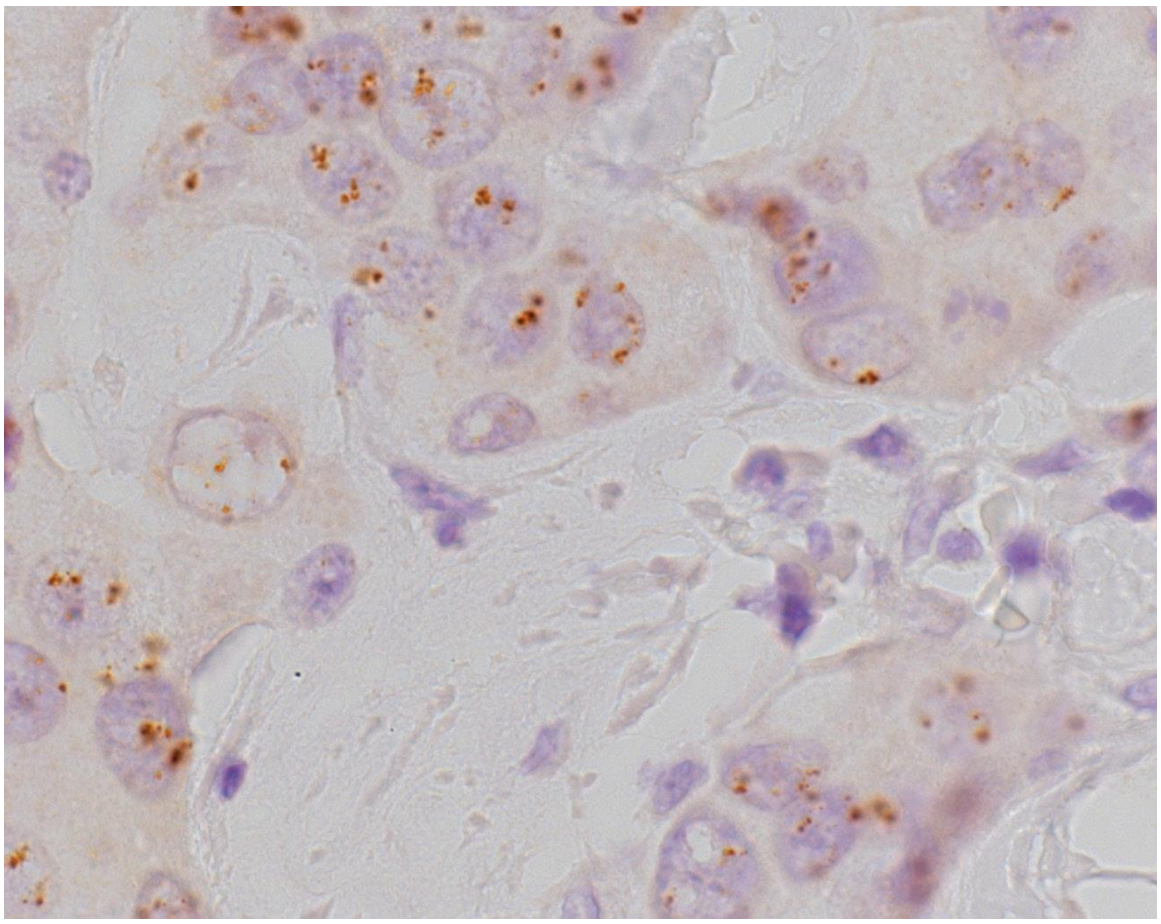
D.

Figure 2: Representative TMA *HER2* staining with IHC, CISH and SISH.

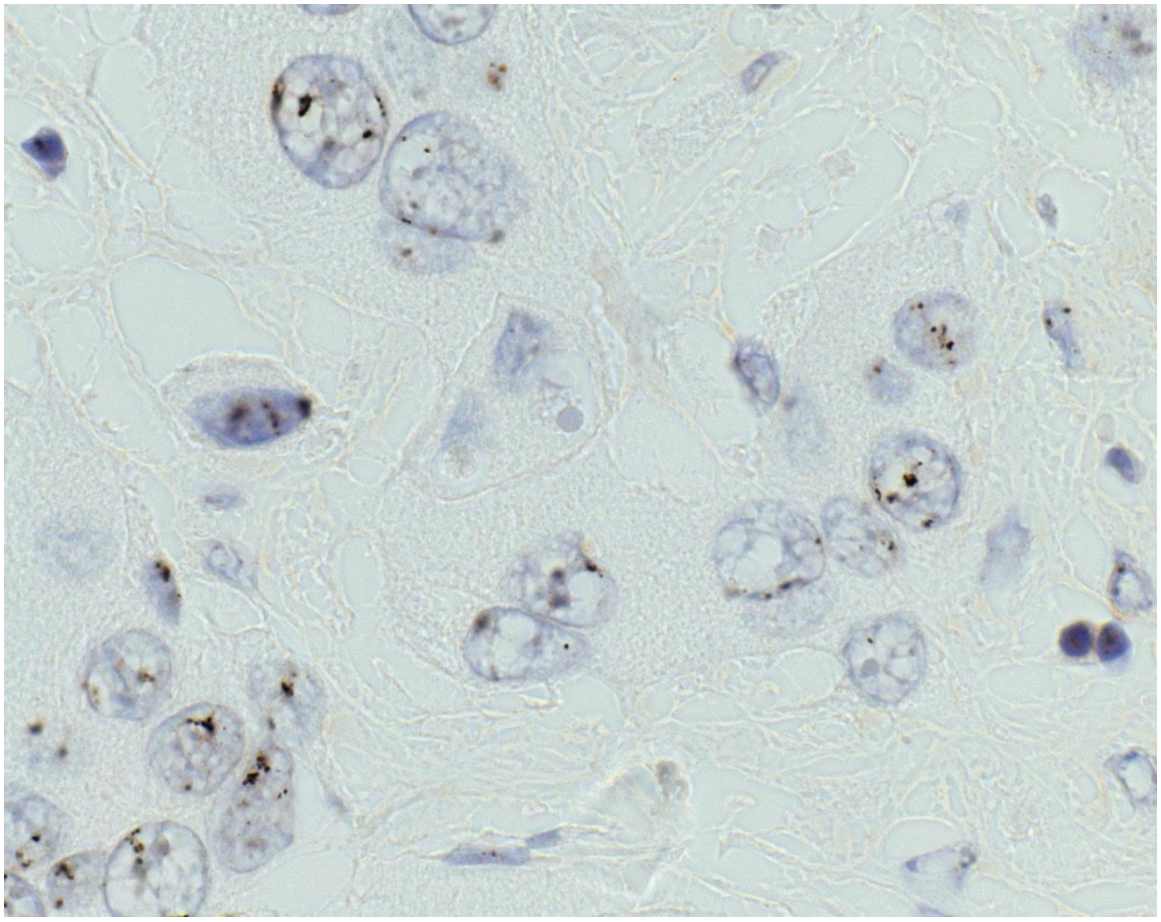
A) IHC 3+ membrane staining (Aperio ScanScope® T2 x20)
B) CISH, Gene amplification (Zeiss AxioImager x630)
C) SISH HER2 gene amplification (Zeiss AxioImager x630)
D) SISH Chr 17 (Zeiss AxioImager x630)



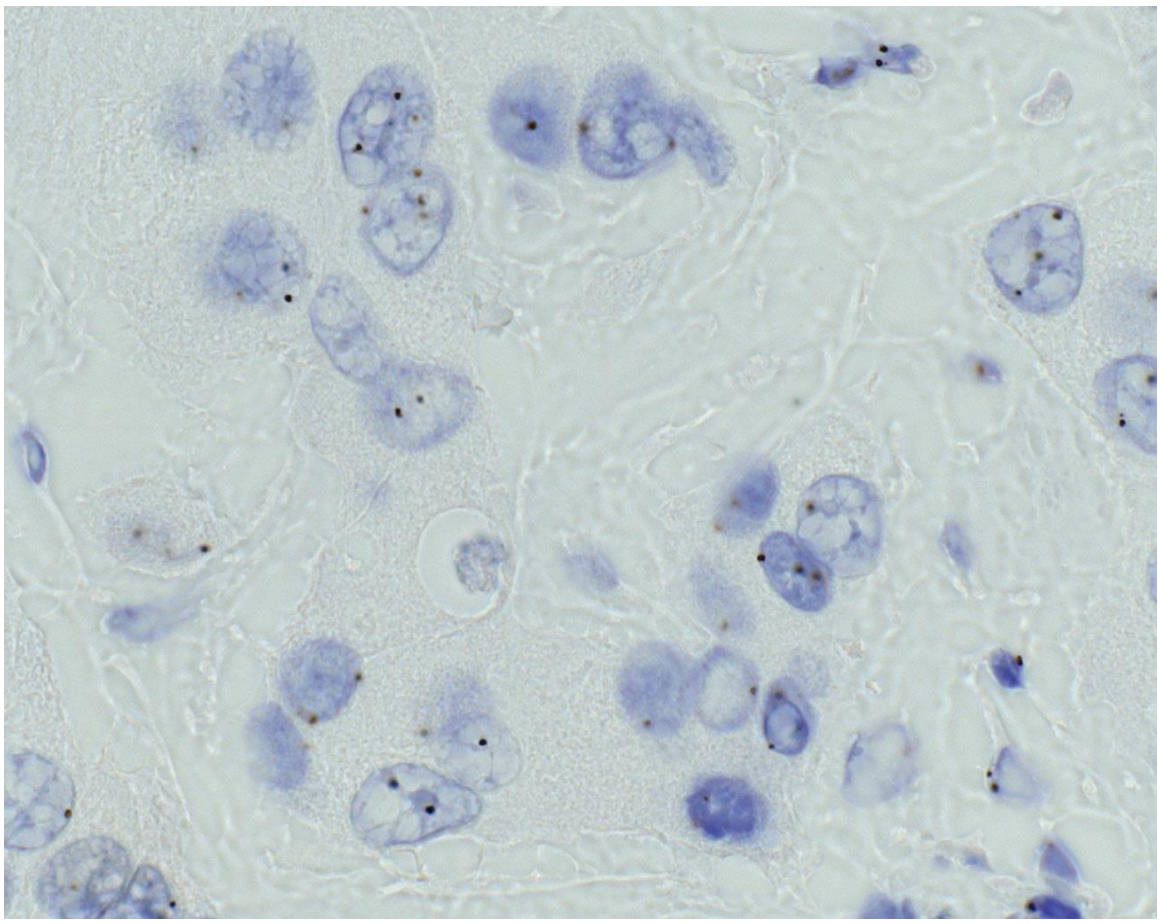
A.



B.



C.



D.

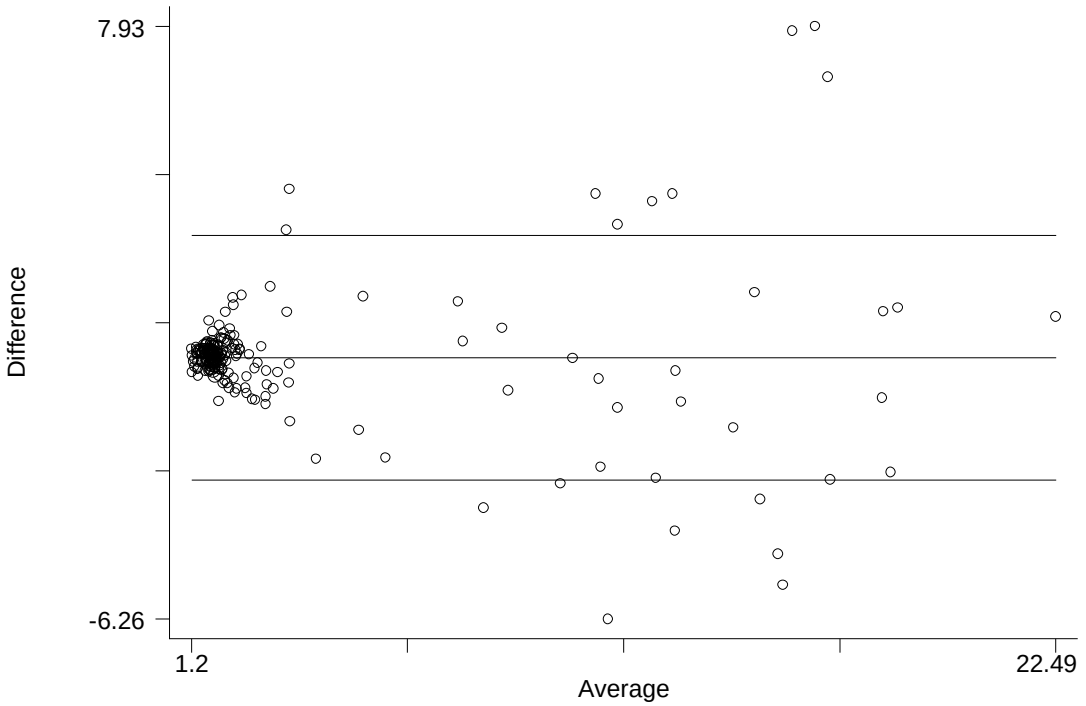


Figure 4: Bland-Altman comparison of CISH count and SISH count.
Limits of agreement (Reference Range for difference): -2.932 to 2.928
Mean difference: -0.002 (CI -0.192 to 0.188)
Range : 1.200 to 22.490
Pitman's Test of difference in variance: r = -0.000, n = 230, p = 0.998

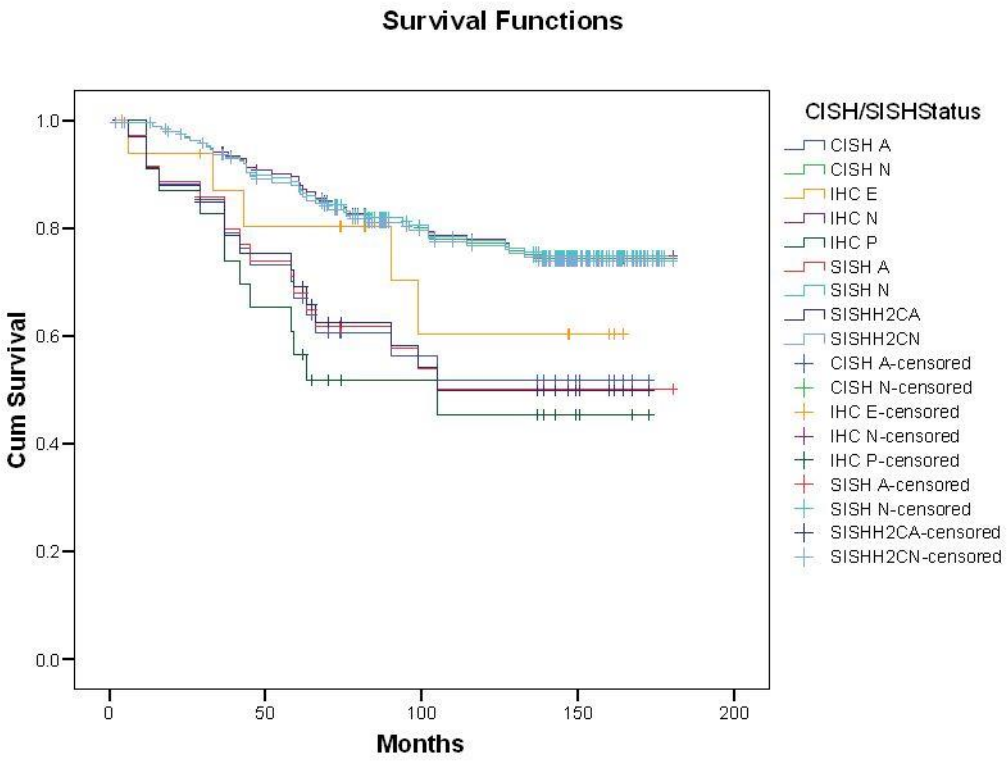


Figure 5: Kaplan-Meier plots for breast cancer specific survival.

Codes: CISH A= low and high amplified cases; CISH N=diploid, polysomy and equivocal cases; SISH A= low and high amplified cases; SISH N=diploid, polysomy and equivocal cases;SISHH2CA= low and high amplified cases utilizing the *HER2*/CHR17 ratio; SISHH2CN= diploid, polysomy and equivocal cases utilizing the *HER2*/CHR17 ratio;IHC E= 2+ IHC staining, IHC N= 0 & 1+ IHC staining and IHC P= 3+ IHC staining.

Chapter 4

WEE, E. J., PETERS, K., NAIR, S. S., HULF, T., STEIN, S., WAGNER, S., BAILEY, P., LEE, S. Y., QU, W. J., BREWSTER, B., FRENCH, J. D., DOBROVIC, A., FRANCIS, G. D., CLARK, S. J. & BROWN, M. A. 2012. Mapping the regulatory sequences controlling 93 breast cancer-associated miRNA genes leads to the identification of two functional promoters of the Hsa-mir-200b cluster, methylation of which is associated with metastasis or hormone receptor status in advanced breast cancer. Oncogene, 31, 4182-95.

4.1.1. Introduction

There are more than 2000 microRNAs (miRNAs) which regulate an estimated 30% of human genes (McGuire et al., 2015). Differential expression of miRNAs have been shown to correlate with metastatic disease in breast cancer and have been associated with lymphatic metastases, however, the expression does not appear to be specific for lymph node metastases and different miRNAs have been associated with metastases in other tumour types (McGuire et al., 2015). The breast cancer database was utilised to select tumours with associated lymph node metastases to perform miRNA analysis and evaluate the regulation of miRNAs.

4.1.2. Results

MicroRNAs (miRNAs) are small non-coding RNAs of B20 nt in length that are capable of modulating gene expression post-transcriptionally. Although miRNAs have been implicated in cancer, including breast cancer, the regulation of miRNA transcription and the role of defects in this process in cancer is not well understood. In this study we have mapped the promoters of 93 breast cancer- associated miRNAs, and then looked for associations between DNA methylation of 15 of these promoters and miRNA expression in breast cancer cells. The miRNA promoters with clearest association between DNA methy- lation and expression included a previously described and a novel promoter of the Hsa-mir-200b cluster. The novel promoter of the Hsa-mir-200b cluster, denoted P2, is located B2 kb upstream of the 50 stemloop and maps within a CpG island. P2 has comparable promoter activity to the previously reported promoter (P1), and is able to drive the expression of miR-200b in its endogenous genomic context. DNA methylation of both P1 and P2 was inversely associated with miR-200b expression in eight out of nine breast cancer cell lines, and in vitro methylation of both promoters repressed their activity in reporter assays. In clinical samples, P1 and P2 were differentially methylated with methylation inversely asso- ciated with miR-200b expression. P1 was hypermethylated in metastatic lymph nodes compared with matched primary breast tumours whereas P2 hypermethylation was associated with loss of either oestrogen receptor or progesterone receptor. Hypomethylation of P2 was associated with gain of HER2 and androgen receptor expression. These data suggest an association between miR-200b regulation and breast cancer subtype and a potential use of DNA methylation of miRNA promoters as a component of a suite of breast cancer biomarkers.

Introduction

Breast cancer is a heterogeneous disease that can be classified on the basis of a number of characteristics including tumour size, histological subtype and grade, oestrogen (ER) and progesterone (PR) receptor and HER2 expression, axillary lymph node (LN) status and expression profile (Sorlie et al., 2001). Some of these features have been associated with disease characteristics and can therefore be used to inform patient management. For example, patients with tumours that test positive for ER and HER2 can be treated with tamoxifen and Herceptin®, respectively, and have a significantly better prognosis than those that test negative for these markers. However, the heterogeneity that exists even within breast cancer subgroups defined by multiple markers means that for the vast majority of breast cancer cases, predicting outcome remains a challenge, and thus additional informative biomarkers are urgently needed.

Breast cancer results from abnormalities in the quality or quantity of certain gene products, including coding and non-coding genes. MicroRNAs (miRNAs) are small non-coding RNAs B20 nt in length that are capable of modulating gene expression post-transcriptionally (Cullen, 2004; Boyd, 2008; Bartel, 2009). MiRNAs can exhibit either tumour suppressor or oncogenic roles by modulating key cellular processes in cell-cycle progression, apoptosis and invasion (Bartels and Tsongalis, 2009; Mirnezami et al., 2009; Visone and Croce, 2009). In several studies, differential miRNA expression has been shown to distinguish normal and breast tumour tissue, breast cancer subtypes, ER, PR and HER2 status, and to predict lymph node status and invasiveness (Iorio et al., 2005; Mattie et al., 2006; Foekens et al., 2008; Yan et al., 2008; Lowery et al., 2009). Together, these studies suggest a potential diagnostic and prognostic use of miRNAs as biomarkers in breast cancer.

Quantitative defects in miRNAs arise through several mechanisms, including aberrant DNA methylation. Human DNA methylation usually occurs at the number 5 carbon of cytosine of a CpG dinucleotide motif. High densities of CpGs, termed CpG islands (CGIs) are usually associated with promoter elements and methyla- tion of which usually leads to gene repression. Aberrant DNA methylation of miRNA genes has been associated with several cancers (Lujambio et al., 2007; Lehmann et al., 2008; Lodygin et al., 2008), suggesting a possible use of miRNA DNA methylation as a prognostic tool. For example,

miR-9-1 and miR-34a are hypermethylated in breast cancer (Lehmann et al., 2008; Lodygin et al., 2008). In addition, the mir-200b cluster (miR- 200b, -200a and -429), has a CGI associated promoter B4 kb upstream of the sequence encoding the mature miRNA (Bracken et al., 2008), and aberrant DNA methylation of this sequence is associated with loss of miR-200 expression in colon (Han et al., 2007), bladder (Wiklund et al., 2011) and pancreatic (Li et al., 2010) cancers. The contribution of miR-200b cluster gene methylation to breast cancer has not yet been reported. Although much is known about the biogenesis and function of miRNAs, relatively little is known about the transcriptional regulation of miRNA genes. To date, a limited number of miRNA promoters have been experimentally characterized and only recently have several miRNA promoter prediction algorithms emerged (Zhou et al., 2007; Fujita and Iba 2008; Linhart et al., 2008; Marson et al., 2008; Ozsolak et al., 2008; Wang et al., 2009). These studies show that miRNA promoters lie anywhere from a few bases upstream of the stemloop to tens of kilobases upstream (Linhart et al., 2008; Marson et al., 2008; Ozsolak et al., 2008; Wang et al., 2009). Furthermore, several miRNAs have multiple promoters (Ozsolak et al., 2008; Wang et al.2009; Monteys et al., 2010).

In this study, 93 miRNAs previously associated with breast cancer, were prioritized for experimental analysis using bioinformatics to look for CGI-associated promoters. The CGI-associated promoters of 15 miRNAs were mapped and methylation determined in a panel of nine breast cancer cell lines. A novel promoter for the miR-200b cluster and its role in regulating miR-200b expression was investigated. The relationship between methylation of this promoter and the previously described miR-200b cluster promoter with miRNA-200b expression and clinical characteristics in breast cancer are described.

Results

Fifty-five miRNAs previously implicated in breast cancer are located within 5 kb of a predicted CGI

To identify candidate promoters for which methylation could be associated with breast cancer development, a list of 93 miRNAs implicated in breast cancer was collated from the literature (Supplementary Table 1). Using CpGPlot and CGI searcher, 55 (59%) of these miRNAs had a predicted CGI within 5 kb upstream of the region encoding their 50 stemloop. The Core Boost_HM promoter prediction algorithm was used to predict regulatory elements controlling the transcription of the 55 miRNA-associated CGIs. This

algorithm uses PolII chip binding, histone modifications and DNA motifs associated with promoters to predict putative transcription start sites (TSS) (Wang et al., 2009). Putative promoters were defined by a CpG promoter prediction cutoff score of 0.5, representing a 90% likelihood of TSS within 500 bp of the predicted region. Figure 1 summarizes the process.

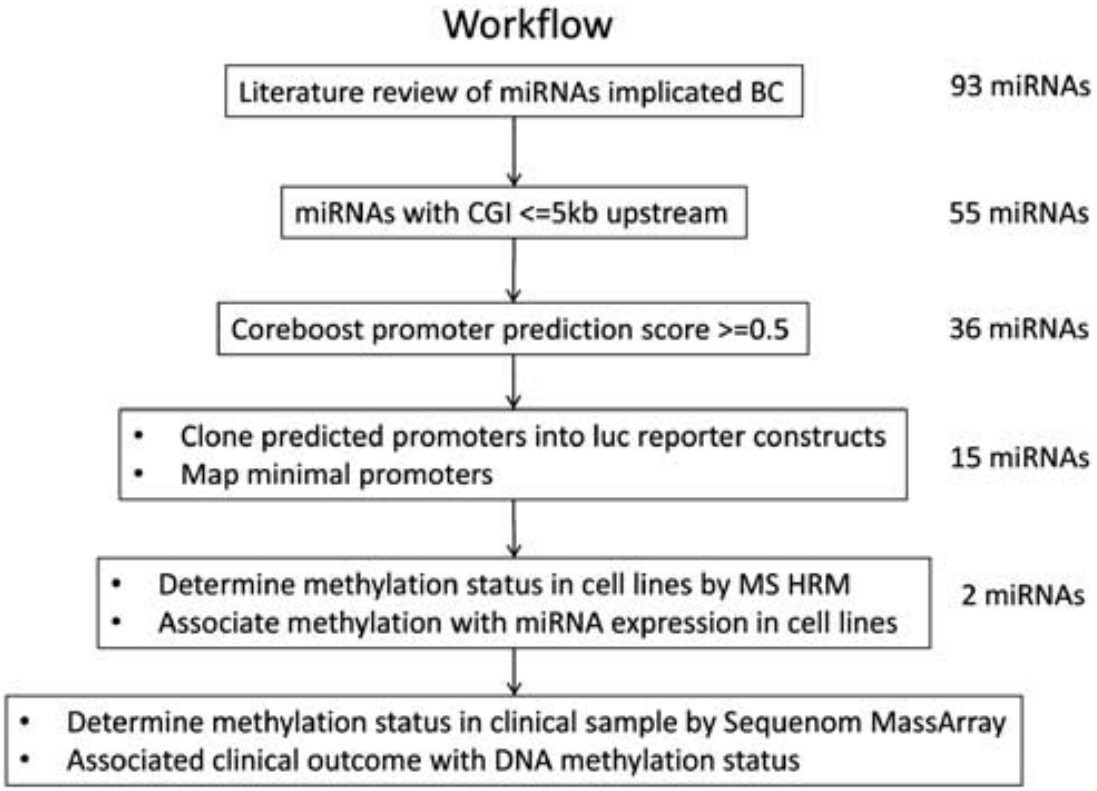


Figure 1: Overview of the miRNA selection approach performed in this study. A literature review of miRNAs implicated in breast cancer was followed by in silico and molecular studies in human breast cancer cell lines. Analysis of methylation in clinical breast cancer specimens was performed on miR-200b.

Experimental validation of predicted novel promoters of 15 miRNAs

To determine if these predicted CpG promoter sequences had experimentally detectable promoter activity, a 600– 1000 bp of genomic sequences around the predicted site was cloned upstream of a luciferase gene and assayed for reporter activities in either MCF7 or MDA-MBD-231 breast cancer cell lines. When a miRNA had more than one predicted promoter, a fragment encompassing each prediction was cloned. Their genomic locations are detailed in Figures 2 and 3 and Supplementary Table 1. The previously described promoters of the miR-17 cluster (Yan et al., 2009) and the miR-200b cluster (Bracken et

al., 2008) were included as positive controls whereas a non-CoreBoost_HM predicted fragment in miR-17 cluster was used to control for background promoter activity. Twenty-two novel promoters from 15 miRNAs exhibited at least fivefold activity compared with the promoter-less pGL3-basic control in at least one cell line (Figures 4–6a). As expected, the previously described promoters of miR-17 and miR-200b clusters had strong promoter activity whereas the non-CoreBoost_HM predicted fragment had no detectable promoter function (Figures 3d and 4a). The 15 miRNAs with experimentally validated promoters are miR-9-1, miR-9-3, miR- 10b, miR-22, miR-124-1, miR-124-2, miR-124-3, the miR-130b cluster, miR-193b, miR-200b cluster, miR-210, miR-320a, miR-335, miR-373 and miR-663. Three promoters were mapped for the miR-124-3 loci; two promoters were mapped for miR-9-1, 22, 124-1, 124- 2, 193b and 200b; and one promoter was mapped for miR-9-3, 10b, 130b, 210, 320a, 335, 373 and 663. To map the minimal promoters regions and to facilitate methylation analysis, promoter fragments of the 15 miRNAs were fine mapped to B300 bp (Figures 2, 3 and 6b).

The miR-200b cluster P2 promoter is sufficient to drive expression of miR-200b

To determine whether the P2 promoter could drive the expression of miR200b in its endogenous genomic context, low miR-200b expressing MDA-MB-231 cells (Gregory et al., 2008), were transfected with a miR-200b minigene spanning the P2, but not the P1, promoter and the sequence corresponding to the mature miR-200. This minigene was generated by replacing the luciferase coding sequence of pGL3-basic with the miR-200b genomic sequence (Figure 6c). The introduction of the miR-200b minigene resulted in an eightfold increase in mature miR-200b expression over the pGL3-basic control (Figure 6c). Deletion of the minimal promoter in the minigene reduced miR-200b expression by 50% (Figure 6c). Collectively, these results indicate that the P2 promoter can regulate miR-200b, and very possibly mir-200a and 429, as a polycistronic primary transcript (Bracken et al., 2008).

The miR-200b cluster P1 and P2 promoters are independent

To address the hypothesis that P1 and P2 promoters function synergistically to enhance expression of the miR-200b cluster, a 2.5-kb fragment encompassing both promoters was cloned upstream of the luciferase gene and assayed for reporter activity in MDA-MB-231 cells, in which only P2 was observed to be functional, and MCF7 cells, in which functional activity was observed for both promoters. As predicted, the P1 þ P2 fragment produced

similar reporter activity to the P2 fragment alone in MDA-MB-231 cells (P ¼ 0.34) (Figure 6a). In contrast, in MCF7 cells, the reporter activity of P1 þ P2 was not significantly greater than the activity of P1 (P ¼ 0.4), but was significantly stronger than P2 (Po0.05) (Figure 6a). Also, the activity of P2 alone and P1 þ P2 were also significantly higher (Po0.05) in MCF7 than in MDA-MB-231 cells (Figure 6a). Taken together, this data suggests that the P1 and P2 promoters function independently.

The miR-200b cluster has multiple TSS

To complement the promoter mapping experiments, attempts were made to map the TSS of P2. Classical 50 RACE PCR was employed to determine the TSS of P2 using RNA from MDA-MB-231 cells transfected with the P2 construct to enrich for P2 derived transcripts. However, repeated attempts with the classical 50 RACE protocol were unsuccessful, consistently producing non- specific smears (data not shown). Successful amplification of template controls indicated that the cDNA synthesis had worked and that this result was more likely to reflect heterogeneity in miR-200b cluster transcripts. An alternative 'PCR walk' approach to mapping the TSS was performed using a single transcript-specific reverse primer and various forward primers toward the 50 end of the cDNA transcript. The longest transcript extended from -3032 bp to -2447 bp upstream of the 50 stemloop as indicated by loss of PCR amplification (Figure 6d). This observation was in agreement with the both the minigene and the luciferase reporter assays. To further test the hypothesis that miR- 200b has multiple TSS, publicly available breast cancer specific RNA-Sequence and RNA PolII–chip data were analysed at the miR-200b loci (Figure 6e). Multiple RNA-sequence peaks were observed along the CGI for T47D and MCF7 cells indicating expression from the CGI. Furthermore, multiple RNA PolII binding signals in MCF7 cells were detected along the associated CGI suggesting multiple TSS. A strong RNA PolII signal overlapping P1 also suggested preferential transcription from P1 in MCF7 (Figure 6e).

Methylation of miR-200b cluster and miR-335 promoters is associated with reduced miRNA expression

DNA methylation of the minimal promoters of the 15 miRNA was assessed by methylation-sensitive high-resolution melt analysis (Wojdacz and Dobrovic, 2007) in a panel of nine breast cancer cell lines. The proximal miR-9-1 promoter was not included as methylation of this promoter had been previously described (Lehmann et al., 2008). The miR-17 cluster

promoter was also excluded because the high density of CG dinucleotides made it unsuitable for methylation-sensitive high- resolution melt analysis. Promoter methylation was then compared with miRNA expression in the same cell lines. Mir-200b cluster and miR-335 promoter methylation were inversely associated with miRNA expression. For miR-200b cluster, eight out of the nine cell lines displayed an inverse association (Figure 7a, Supplementary Figure 2). Although only MCF7 highly expressed miR-335, MCF7 also had the lowest methylation compared with the remaining eight, which were fully methylated and had minimal miR-335 expression (Supplementary Figure 1A). However, since the inverse association was stronger in miR-200b, P1 and P2 promoters represented better candidates for further analysis. In contrast, the miR-210 and miR-320a promoters were unmethylated in all nine cell lines although DNA methylation was not associated with miRNA expression for miR-9, miR-10b, miR-124, miR- 373 and miR-663 (Supplementary Figure 1).

The minimal miR-200b cluster promoters are regulated by DNA methylation

The novel P2 promoter had comparable activity to P1 in MCF7 cells, but unlike P1, was functional in both cell lines tested (Figure 6a). The minimal P2 promoter maps to -2228/-1993 bp upstream of the miR-200b 50 stemloop (Figure 6b). To confirm that DNA methylation directly repressed promoter activity, P1 and P2 were cloned into a CpG-free reporter construct (Klug and Rehli, 2006) and in vitro methylated by SssI DNA methylase. Methylated P1 and P2 constructs displayed a significant reduction in promoter activity, compared with their mock methylated constructs when transfected into T47D cells, in which both promoters are endogenously methylated and functional (Figure 7b). This suggests that DNA methylation represses miR200b cluster promoter activity.

miR-200b P1 and P2 promoters are differentially methylated in primary breast tumours

To study DNA methylation of the miR-200b promoters, Sequenom MassArray was performed on Grade 3 FFPE clinical samples. In all cases, P1 and P2 were differen- tially methylated in both tumours and lymph nodes (Figures 8a and b). In addition, P1, but not P2, was hypermethylated in lymph nodes compared with matched primary tumours (Figures 8c and d). To determine if hypermethylation was associated with expression of the miR-200b cluster in primary tumours, qPCR for miR-200b was performed on tumour samples from which RNA was available. P1 hypermethylation was associated with loss of miR-200b expression in seven out of nine samples (Supplementary Figure 3A) whereas

P2 was found to be associated with loss of miR- 200b expression in six out of seven samples tested (Supplementary Figure 3B). These suggested that hypermethylation of miR-200b cluster promoters could regulate miRNA expression in tumours.

Methylation of the miR-200b P2 promoter is associated with ER, PR, HER2 and androgen receptor expression in primary breast tumours

To ascertain whether DNA methylation of the miR-200b cluster promoters is associated with expression of routinely used breast cancer biomarkers, ER, PR and HER2, methylation was assessed in patients positive and negative for expression of these receptors. Methylation of P2, but not P1, was significantly higher in tumours that were ER or PR negative (Figures 9a and b, respectively). Hypermethylation of P2 was also associated with HER2 positivity (Figure 9c). Androgen receptor, a potential breast cancer biomarker (Hu et al., 2011) and regulator of the miR-200 family, (Xu et al., 2010; Waltering et al., 2011) was also associated with hypermethylation of P2 (Figure 9d). Although the mir- 200b cluster is involved in metastasis, which in turn affects prognosis, no evidence of an association between DNA methylation and survival was found.

Discussion

Transcriptional regulation of miRNA genes is poorly understood and only a few miRNA promoters have been reported. A comprehensive understanding of miRNA promoters is a prerequisite for their use as genetic or epigenetic biomarkers. In this report, novel CGI-associated miRNA promoters were mapped and ana- lysed for associations between DNA methylation and miRNA expression. In all, 59% of the miRNAs examined were associated with a CGI within 5 kb upstream, similar to the estimated proportion of CGI- associated coding genes and was consistent with previous estimations for miRNA promoters (Ozsolak et al., 2008; Corcoran et al., 2009). Twenty-two novel promoters were identified and shown to be active in reporter assays. MiR-10b had a previously described promoter (Ma et al., 2007; Zhou et al., 2007) immediately upstream of the mature miRNA sequence (Figure 2). However, we could not detect promoter activity for this fragment in the breast cancer cells tested (Figure 4c). We were also unable to detect any activity in fragments encompassing Core- Boost_HM predicted promoter regions for miR-125a in the cell lines used (Figures 3 and 4i). A likely explanation was that neither cell line expressed miR-125a.

In all, 7 of 15 miRNAs had two or more promoters in close proximity, usually at either ends of the associated CGI. Although it was not clear how the multiple promoters function in

regulating their miRNA genes or why the promoters were usually at either end of the CGI, it was evident that regulation of miRNAs is a complex process. The miR-200b cluster (miR-200a, miR-200b and miR-429) is an example of a miRNA with promoters at either end of the CGI. The P1 promoter, located at the distal end of the CGI, was predicted (Bracken et al., 2008) based on the presence of a 50 EST, the presence of E-Box motifs and the presence of a CGI, which is commonly associated with promoters of coding genes. Further, a 7.5-kb primary transcript of the miR-200b cluster was described using a 'PCR walk' approach and P1 promoter activity was demonstrated using a luciferase reporter assay. Using a similar approach, a novel promoter, P2, is described here. P2 was predicted B2.5 Kb downstream of P1 (Figure 3) by the CoreBoost_HM promoter prediction algorithm (Wang et al., 2009), which utilizes empirical data such as ESTs, RNA PolII binding and histone modification profiles in addition to DNA motifs associated with core promoters to accurately predict active promoter sites. We demonstrate that the P2 promoter has an activity similar to that of the P1 promoter (Figure 6a), is functional in breast cancer cell lines (Figures 6a and 7b) and is able to drive the expression of miR-200b in its endogenous genomic context (Figure 6c). Thus, the P2 promoter is likely to be important in the regulation of the miR-200b cluster. Deletion of the P2 minimal promoter also reduced miR-200b levels by 50% (Figure 6c), and may indicate multiple TSS as previously suggested (Wiklund et al., 2011). In addition, DNA methylation of both miR-200b promoters repressed miR-200b expression in eight out of nine breast cancer cell lines studied (Figure 7), suggesting regulation by DNA methylation. However, the precise role of P1 and P2 in regulating the cluster is unclear. In our reporter assays, P1 and P2 seemed to function independently (Figure 6a). In clinical samples, DNA methylation at P1 was also different compared with P2 in both tumour and lymph node metastases (Figures 8a and b), thus supporting the hypothesis that the two promoters have different regulatory roles. This hypothesis is supported by other studies in bladder cancer cells, where a region encompassing P2, but not P1, was unmethylated and expressed high levels of miR-200b (Wiklund et al., 2011). Taken together, the evidence suggests that the P1 and P2 transcripts are regulated by different mechanisms and this could in turn have a role in regulating metastasis.

Of the eight cell lines studied, MCF7 did not show an inverse association between methylation and miRNA expression (Figure 7a). A minority of patients also did display a reciprocal relationship between promoter methylation and miRNA levels (Supplementary Figure 3). In previous studies (Han et al., 2007; Wiklund et al., 2011), miRNA repression

by DNA methylation is usually accompanied by histone modifications associated with gene silencing. Thus, other mechanisms including chromatin remodelling or post-transcriptional regulatory events may account for this inconsistency. Perhaps these repressive histone marks were absent in these cases thus resulting in open chromatin that was readily expressed.

In this study we describe, for the first time, differential methylation of the P1 and P2 region of the miR-200b cluster in breast cancer. The differential methylation is functional is evidenced by our observations that DNA methylation is inversely associated with miR-200b expression in both breast cancer cell lines (Figure 7) and clinical samples (Supplementary Figure 3). These are consistent with the previously reported tumour suppressive role of miR-200b (Korpal and Kang, 2008). They are also consistent with previous reports of aberrant DNA methylation of the miR-200b cluster proximal CGI, containing both P1 and P2, in colon, bladder and pancreatic cancers (Han et al., 2007; Li et al., 2010; Wiklund et al., 2011).

Loss of ER and PR expression was also associated with DNA methylation at P2 in breast tumours (Figures 9a and b). Patients with tumours that express these receptors often have a better prognosis because they respond well to treatments such as Tamoxifen. We hypothesize that methylation at P2, is likely to be associated with a lower level of miRNA expression, resulting in a more aggressive tumour (Korpal and Kang, 2008) that is unresponsive to these therapies and generally associated with poor prognosis. Using publicly available ER Chip-sequence data (Schmidt et al., 2010), ER bound to a putative ER response element just downstream of P1 upon ER stimulation in MCF7s (Figure 6e). In a microarray study (Klinge, 2009), miR- 200a and miR-200b were significantly upregulated in MCF7 after 6 h of E2 induction. However, in a similar independent study, miR-200a and 200c were found to be significantly downregulated after 48 h of E2 induction (Maillot et al., 2009). Although the studies seemed to have conflicting conclusions, they do suggest a possible regulatory mechanism between ER and the miR-200 family.

A double negative feedback regulatory relationship between the miR-200 family and ZEB1 (Bracken et al., 2008; Burk et al., 2008) has been shown to regulate the delicate balance between mesenchymal and epithelial cellular states. Based on this data, we propose that miR- 200b is repressed in the early stages of tumourigenesis in order to promoter EMT and thus the spread of the tumour, followed by later induction of miR-200b to promote mesenchymal–epithelial transition and thus establishment of the tumour cells at a distant site (for example, lymph node). Our data is consistent with this as we show only P1 was

hypermethylated in matched lymph nodes compared with their primary tumours. This coupled with miRNA repression, suggests a DNA methylation mechanism for EMT initiation in addition to the previously described TGFB/ZEB pathway. At P2, no differential methylation between primary tumours and matched lymph node and thus possibly maintaining base levels of miR-200 is consistent with the mesenchymal–epithelial transition observed in mouse models. Metastatic murine breast cancer cells expressing low levels of miR-200 were able to invade distant tissue but unable to colonize. However, when miR-200 was over- expressed, these cells could form macroscopic tumours at distant sites (Dykxhoorn et al., 2009). Further support for this model comes from studies in the bladder cancer (Wiklund et al., 2011) where hypomethylation of the P2 region was sufficient for miR-200b cluster expression. This hypomethylation could also possibly account for the elevated levels of the miR-200 family, observed in other cancers (Hiroki et al., 2010; Li et al., 2010; Lee et al., 2011).

Collectively, the evidence presented here indicates that miR-200b cluster regulation is complex and is regulated transcriptionally by at least two distinct promoters that are sensitive to DNA methylation. The novel P2 promoter functions independently of P1 and can drive the expression of miR-200b. However, the precise roles of P1 or P2 and under what conditions they are utilized is still not clear and will require further examination.

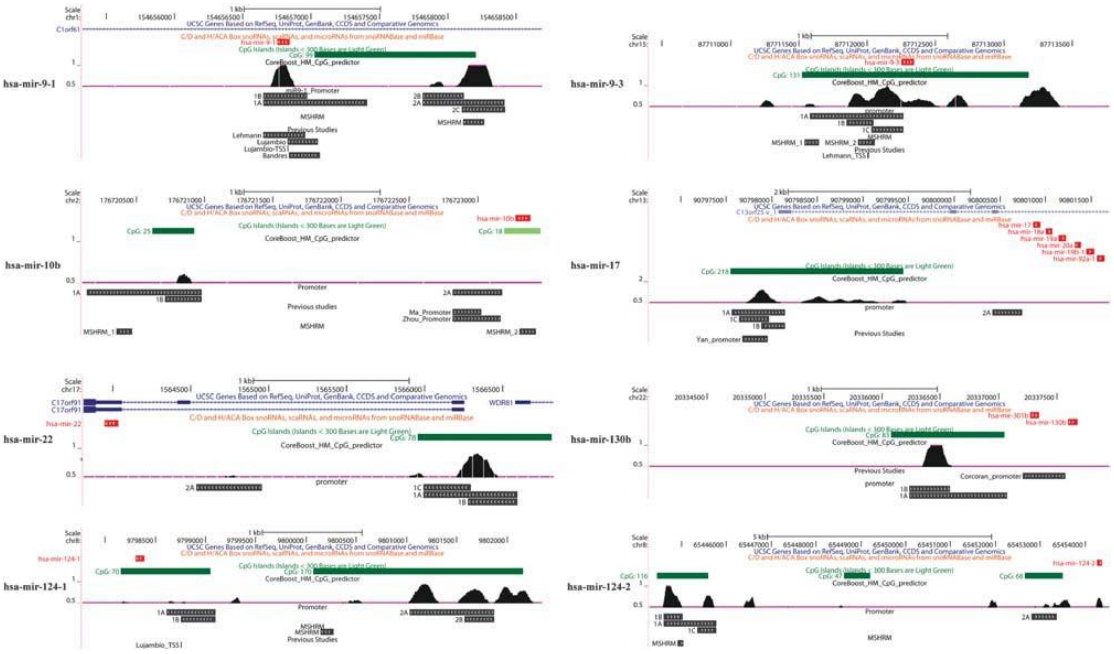


Figure 2 UCSC screenshots of miRNA candidates and their associated genomic features. Bars representing miRNAs are shown in red, CGIs in green, promoter and methylation-sensitive high-resolution melt analysis fragments in black. Annotated genes are marked in

blue. Orientation of genes and fragments are indicated by directional arrows. CoreBoost_HM promoter predictions are shown as black peaks.

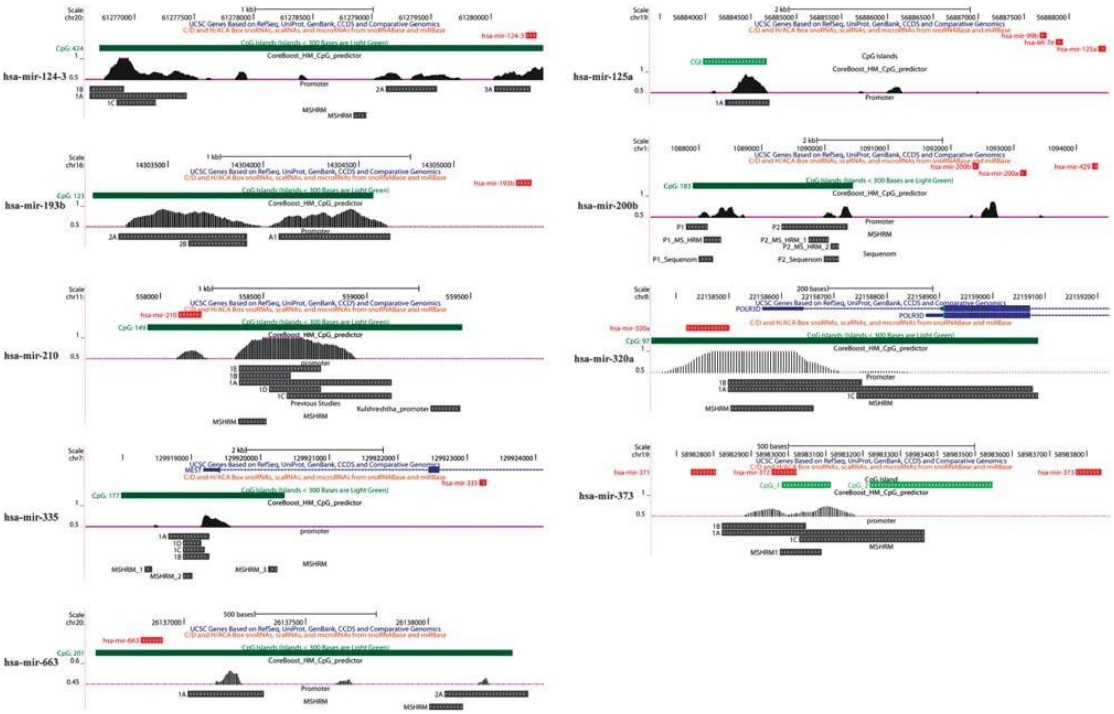


Figure 3 UCSC screenshots of miRNA candidates and their associated genomic features. Bars representing miRNAs are shown in red, CGIs in green, promoter and methylation-sensitive high-resolution melt analysis fragments in black. Annotated genes are marked in blue. Orientation of genes and fragments are indicated by directional arrows. CoreBoost_HM promoter predictions are shown as black peaks.

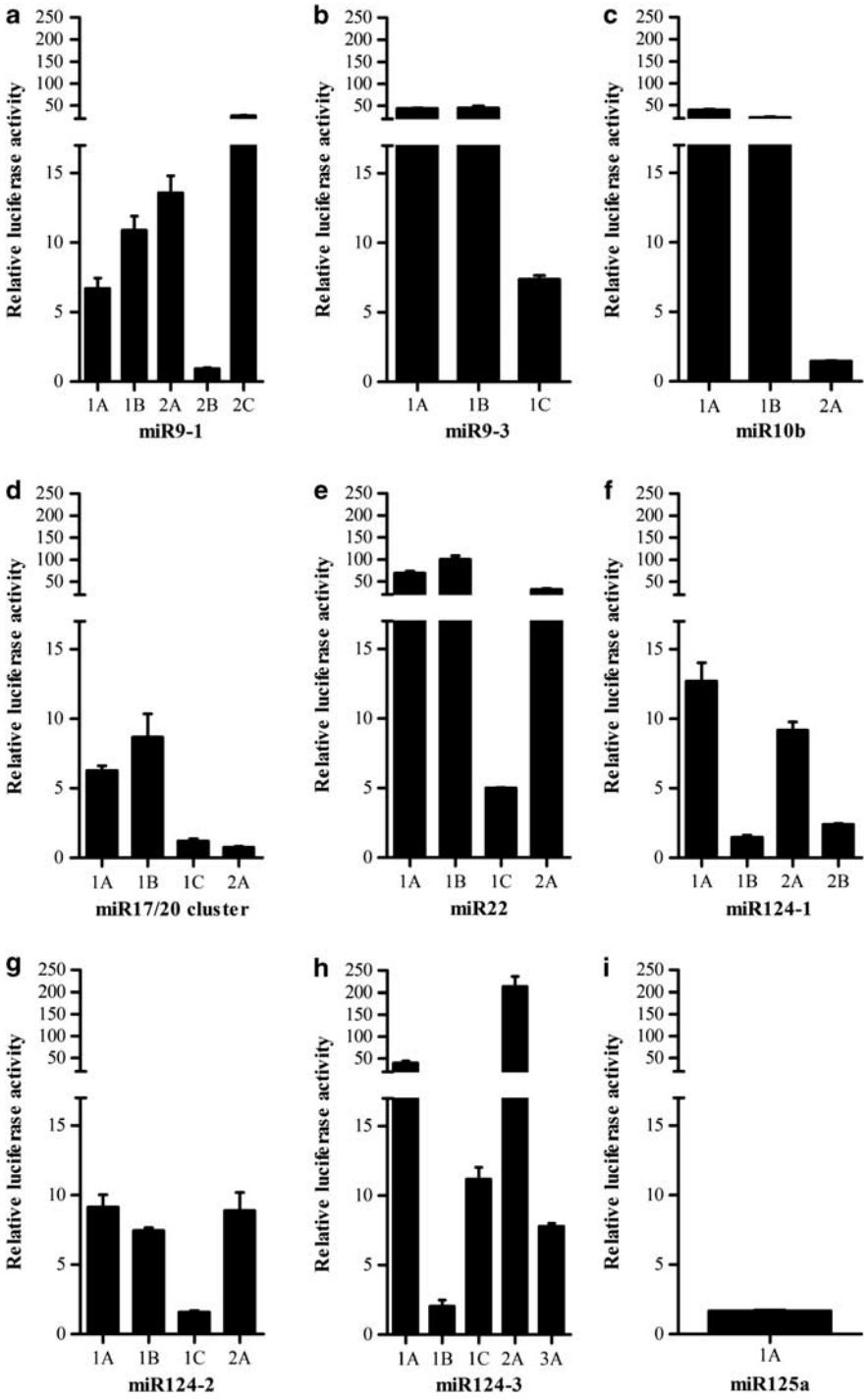


Figure 4 (a–i) Promoter activities of miRNA candidates. miRNA promoter activity in cells expressed in RLU±the s.e.m. Data were generated from three independent experiments. Promoter fragments are labelled 1, 2 or 3 and A, B or C indicates the sub-fragment of that respective promoter.

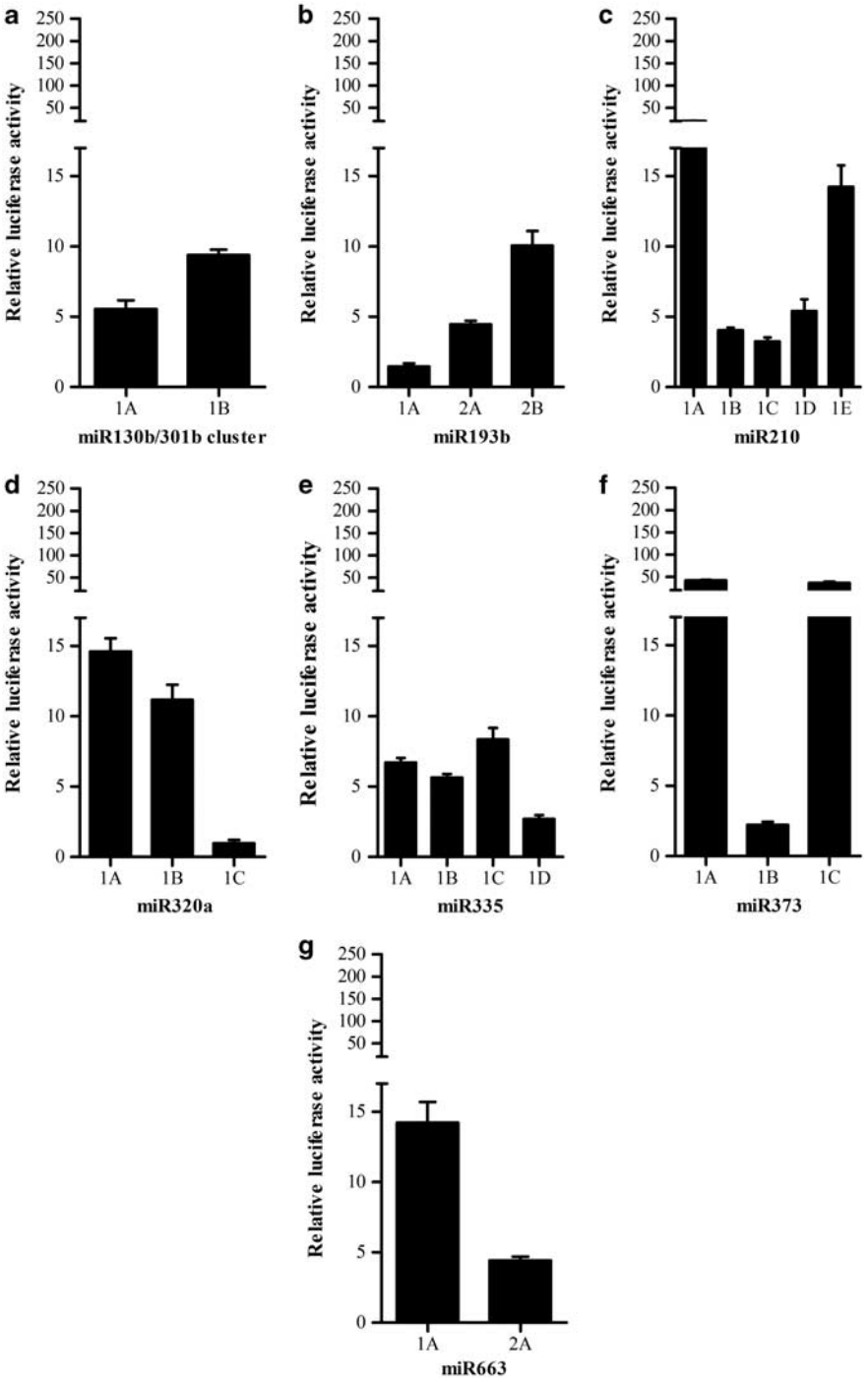


Figure 5 (a–g) Promoter activities of miRNA candidates. miRNA promoter activity in cells expressed in RLU±the s.e.m. Data were generated from three independent experiments. Promoter fragments are labelled 1, 2 or 3 and A, B or C indicates the sub-fragment of that respective promoter.

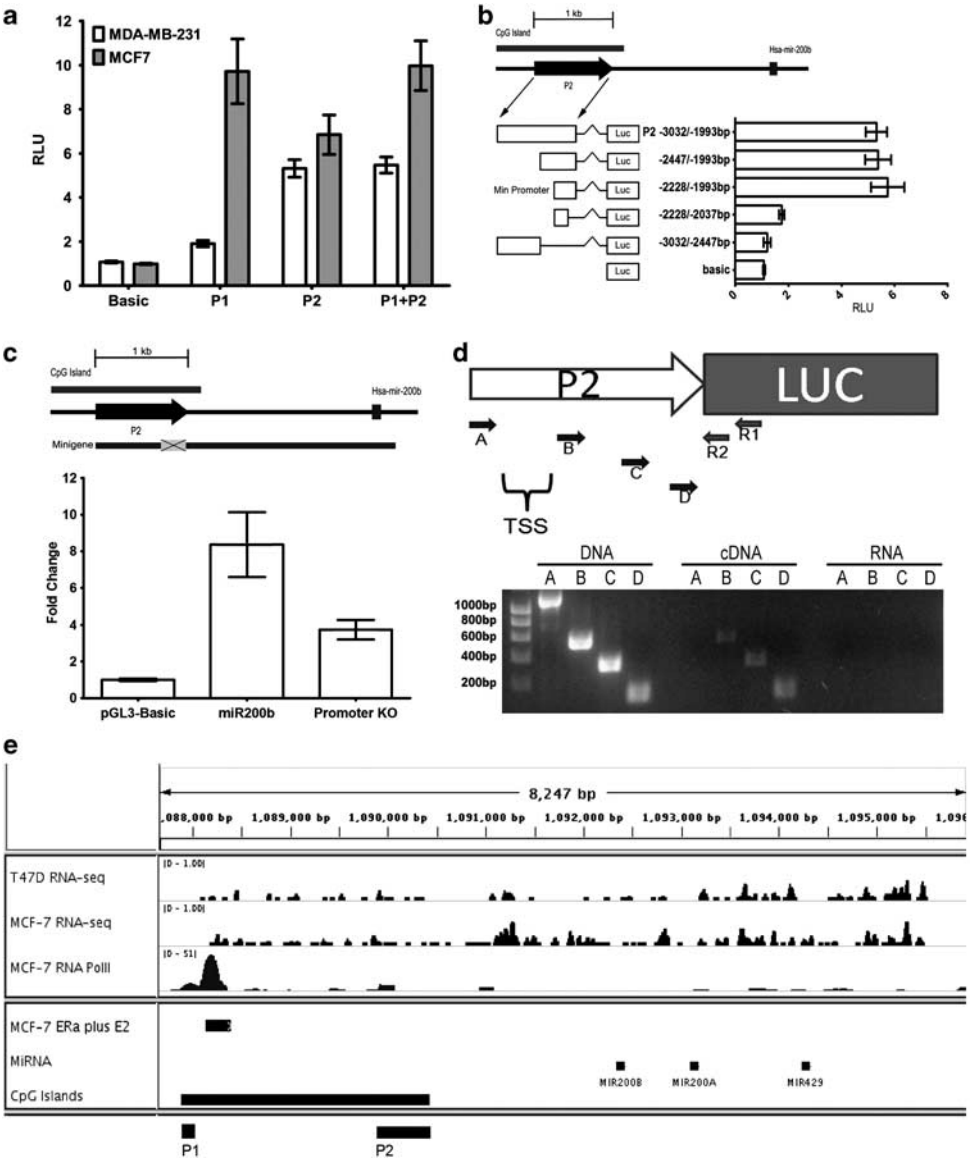


Figure 6 miR-200b cluster has two functional promoters. (a) Promoter activities of miR-200b P1 and P2 in MCF7 and MDA-MB-231 cells. Reporter activity is expressed in RLU±the s.e.m. Data were generated from three independent experiments. White bars represent activity in MDA-MB-231 cells whereas grey bars represent activity in MCF7 cells. (b) Luciferase activities of various miR- 200b P2 50 and 30 truncations. Graphical representation of the various reporter fragments in relation to their genomic locations and their associated reporter activities expressed in RLU. Error bars represent s.e. of three separate experiments. (c) Top: relative location of the minigene to P2, Hsa-mir-200b and the CGI. Grey box represents the region deleted in the minigene promoter KO construct. Bottom: MDA-MB-231 cells were transfected with pGL3-basic, miR200b minigene or miR200b minigene promoter KO plasmids and assayed for mature miR-200b expression by TaqMan real-time PCR±s.e. (d) Top: schematic of the miR-200b P2 luciferase construct.

The white arrow represents the P2 promoter followed by the luciferase gene. Primers R1, R2, A to D are represented small arrows. Bottom: DNA agarose gel of the 50 PCR walk using the R2 primer with either Primer A, B, C or D, as labelled above each lane.
(e) RNA-sequence profiles in T47D and MCF7 and RNA Polymerase II (RNA PolII) chip profile of MCF7 at the miR-200b locus. Peaks in RNA-sequence tracks represent expression detected at that region. Peaks represent RNA PolII binding in RNA PolII track. Horizontal bars indicate the location of the miR-200b cluster, CGI and oestrogen response element.

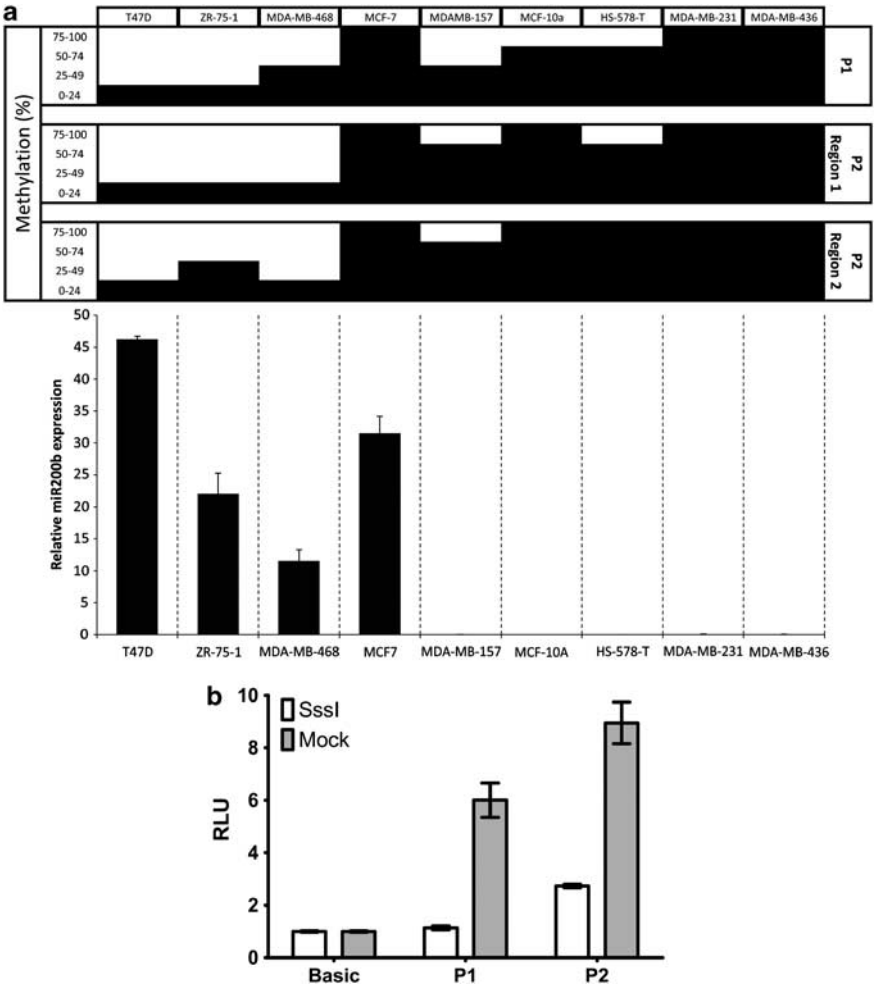


Figure 7 DNA methylation represses miR-200b P1 and P2 activity in breast cancer cells. (a) DNA methylation and miR-200b expression levels in a panel of nine breast cancer cell lines. Top: black bars represent the percentage methylation. Bottom: miR-200b expression was assessed by qPCR. Expression is shown relative to RNU6B and bars represent the mean±s.e. of two independent experiments. (b) Reporter activity of miR-200b P1 and P2 methylated by SssI DNA methylase (white) compared with mock methylated plasmids (grey)±s.d. of two separate experiments.

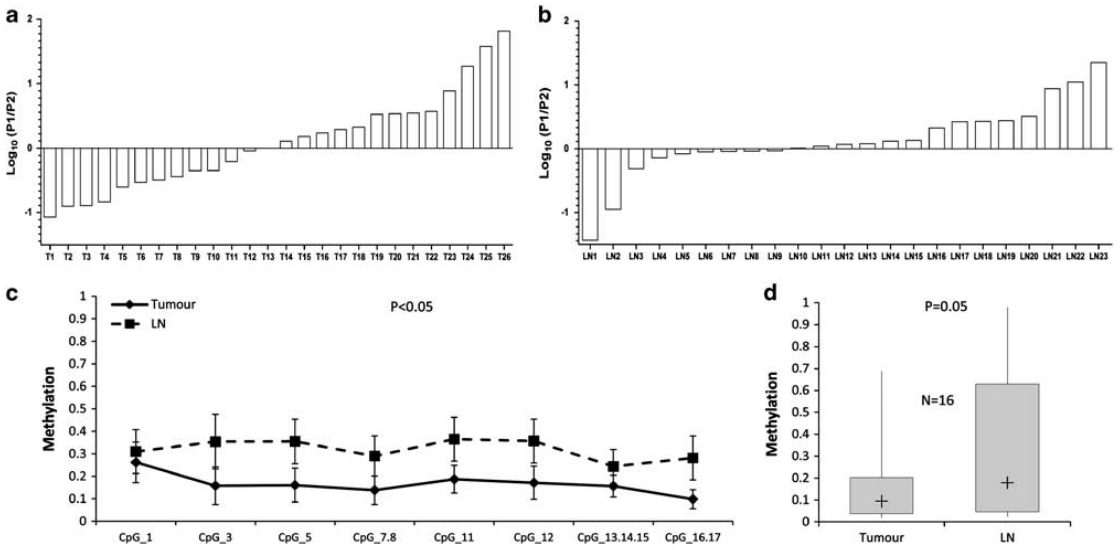


Figure 8 Differential methylation of P1 and P2 in clinical samples. (a) Log10 ratios of P1 to P2 in 26 primary tumours. (b) Log10 ratios of P1 to P2 in 23 lymph nodes (LN). Positive values: P1>P2; negative values: P1oP2. Graph heights represent magnitude of difference in methylation between P1 and P2. (c) Mean methylation profile in matched tumours and LN with horizontal s.e. bars for individual CpG units. t-Test P-value as indicated. (d) Box plot of the average methylation in tumours and LN. (þ): median, box: 25–75 percentile, whiskers: max/min, N: sample size, Mann–Whitley P-values as indicated.

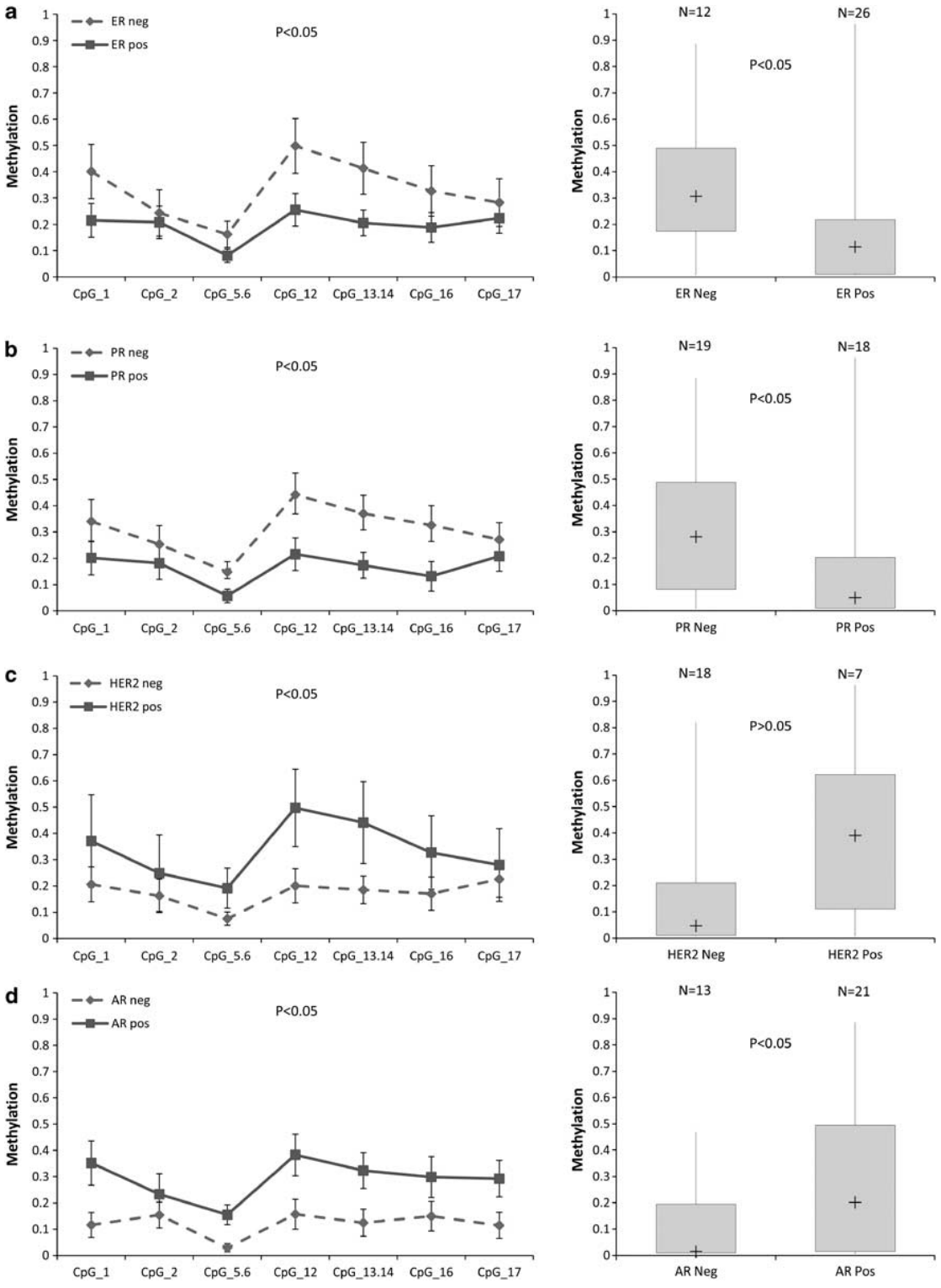


Figure 9 P2 methylation is associated with ER, PR, HER2 and AR receptor status. (a) Methylation status of ER positive (pos) and ER negative (neg) cohorts. (b) Methylation status of PR pos and PR neg cohorts. (c) Methylation status of HER2 pos and HER2 neg cohorts. (d) Methylation status of AR pos and AR neg cohorts. Left: mean methylation profile with horizontal s.e. bars for individual CpG units. t-Test P-value as indicated. Right:

box plot of the average methylation in tumours and LN. (þ): median, box: 25–75 percentile, whiskers: max/min, N: sample size, Mann–Whitley P-values as indicated.

Materials and methods

Bioinformatics

A list of 93 miRNAs implicated in breast cancer was generated by literature review. Genomic sequences þ 5 and -1 kb of the
50 stemloop of each miRNA were analysed for CGIs using CpGPLot and CpG Island Searcher. Putative promoters are defined by a CoreBoost_HM score of at least 0.5 (Wang et al., 2009) within this 6 kb window. Initial 600–1000 bp fragments overlapping the predicted sites were cloned and assayed for promoter activity as described later. This process is illustrated in Figure 1.

RNA Polymerase II Chip-sequence data mapped to human genome HG18 was obtained from the National Center for Biotechnology Gene Expression Omnibus, GEO accession number GSE14664. RNA-sequence data (Wang et al., 2008) mapped to human genome HG18. RNA Polymerase II and RNA-sequence data were visualized on integrative genome viewer using the data ranges indicated.

Cell culture

Breast cancer cell lines MDAMB157, MDAMB231, MDAMB436, MDAMB468, MCF7, T47D, ZR75-1, Hs578T and BT549 were obtained from American Type Culture Collection (ATCC, Manassas, VA, USA) and cultured according to the manufacturer's recommendations.

Patient samples

Human breast tumours and matching lymph node metastases were collected from 56 patients, as approved by local Human Ethics committees, who underwent surgical resection and did not undergo preoperative radiochemotherapy at Princess Alexandra Hospital between 1988 and 2000. All patients were female aged from 30 to 94 years old, with a median age of 56 years. ER, PR and HER2 receptor status of each patient were determined by a qualified pathologist. Details are provided in the Supplementary Information.

DNA extractions and purifications

Genomic DNA from cell lines was extracted using the NucleoSpin Tissue Prep kit (Macherey-Nagel, Germany) according to the manufacturer's instructions. Plasmid DNA was purified using the Miniprep Kit (Qiagen, Doncaster, VIC, Australia). For human tumour samples, four FFPE tumour- rich tissue cores (1 x 0.6 mm) were crushed and digested with proteinase K at 55 1C for 2 days. Genomic DNA was purified using the PureGene kit (Qiagen) according to the manufac- turer's instructions.

Generation of plasmid constructs

All promoter reporter constructs were cloned into pGL3-Basic (Promega, Sydney, NSW, Australia) unless otherwise specified. For the in vitro methylation plasmids, P1 and P2 fragments were cloned into a CpG-free luciferase reporter construct pCpG-basic (Klug and Rehli, 2006; a gift from Klug and Rehli). PCR was performed using KapaHiFi polymerase (Kapa Biosystems, Woburn, MA, USA). All constructs were confirmed by sequencing. All primers used and cloning details are provided in the Supplementary.

Transfections and reporter assays

All transfections used a 3-ml: 1 mg ratio of Fugene (Roche, Castle Hill, NSW, Australia) tranfection reagent to DNA. For luciferase assays, either MCF7 or MDA-MB-231cells were co- transfected with 400 ng of promoter construct and 10 ng of RL-TK plasmid (Promega) as a transfection control and harvested and assayed for reporter activity after 48 h. The Dual-Glo luciferase Assay kit (Promega) was used as recommended by the manufacturer. Firefly luciferase levels were normalized to Renilla luciferase levels and expressed relative to pGL3-basic levels (RLU). Statistical analysis was performed using unpaired two-tailed t-test.

For minigene experiments, MDA-MB-231 cells were grown to 60–70% confluence in 6-well plates, and transfected with 1 mg of DNA and harvested after 72 h.

Identification of TSS of Hsa-mir-200b

Total RNA from MCF7 transfected with the P2 luciferase reporter construct was extracted using Trizol (Invitrogen) and DNaseI (NEB, Ipswich, MA, USA) treated. First strand cDNA was reverse transcribed using SuperSciptIII (Invitro- gen) using a luciferase specific primer, R1, at 50 1C. This then served as a template for PCR amplification. PCR 'walking' towards the 50 end was performed using primers A to D with R2. All PCR products were

visualized on a 1% agarose gel. Primer sequences are given in Supplementary Table 3. Classical 50 RACE was perform as previously described (2005).

Quantitation of miRNAs

Total RNA was extracted from cell lines using Trizol (Invitrogen). RNA from clinical samples was extracted using miRNeasy kit (Qiagen). For miR200b and miR335 experiments, cDNA was made from total RNA using TaqMan MicroRNA Reverse Transcription Kit (Applied Biosystems, Mulgrave, VIC, Australia) with both reverse transcription miRNA and RNU6B (loading control) primers in the same reaction. Real-time PCR was performed using the TaqMan microRNA Assay (Applied Biosystems) according to the manufacturer's instructions. For all other miRNAs, the Qiagen miScript PCR system for miRNA quantification was used with the RNU6B loading control. Changes in expression levels were calculated using DDCt method (Livak and Schmittgen, 2001).

Bisulfite modification and methylation-sensitive high-resolution melt analysis

2 mg of DNA extracted from cell lines was subjected to bisulfite modification with MethylEasy Xceed kit (Human Genetic Signatures, Randwick, NSW, Australia) according to manufacturer's instructions. PCR amplification and methylation- sensitive high-resolution melt analysis (Wojdacz and Dobrovic, 2007) was performed in duplicate on the RotorGene Q (Qiagen). Primers were designed according to the principles outlined (Wojdacz and Dobrovic, 2007) to control for PCR bias and are shown in Supplementary Tables 4 and 5. PCR conditions are provided in the Supplementary. Bisulfite treated CpGenome Universal Methylated DNA (Chemicon, Milli- pore, Kilsyth, VIC, Australia) and DNA from the appropriate cell lines were used as positive/methylated and negative/unmethylated controls, respectively. WGA DNA made with the GenomiPhi kit (Amersham GE Healthcare, Rydalmere, NSW, Australia) was used as unmethylated controls for miR335 and miR663. Included in the analysis of each region, controls were mixed in 25, 50 and 75% methylated to unmethylated template ratios.

In vitro methylation of plasmid DNA

DNA was methylated using SssI (NEB) as previously described (Klug and Rehli, 2006). Briefly, plasmids were incubated with SssI (2.5 U/mg) with 160 mM S-adenosylmethio-nine at 37 1C for 4 h and supplemented with an additional 160 mM of S-adenosylmethionine for another 4 h at 37 1C. Mock methylated plasmids controls were

treated similarly but without enzyme. Plasmids were recovered by phenol/chroloform, followed by ethanol precipitation, transfected into T47D cells and luciferase assays performed.

Sequenom MassArray

Genomic DNA from clinical samples was bisulfite converted with EZ-96 DNA methylation kit (Zymo Research, Irvine, CA, USA). Methylation levels in clinical samples were determined using Sequenom MassArray, performed according to manufacturer's recommendations for T-cleavage chemistry protocol and analysed between a 1640 and 7000 mass window (Coolen et al., 2007). Average methylation of each patient is defined as the average percent methylation of all CpG units in each amplicon. Average methylation of each CpG cluster (or profile) is defined as the average percent methylation of the cohort for that specific CpG cluster. In all, 0–100% methylation are represented by 0.0 to 1.0. Primer sequences are provided in Supplementary Table 4.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

This project was funded by grants from the National Breast Cancer Foundation (NBCF), Cancer Australia and National Health and Medical Research Council (NHMRC). We are grateful to Dr Maja Klug for generously providing the CpG- Less reporter plasmid.

References

(2005). Rapid amplification of 50 complementary DNA ends (50 RACE). Nat Methods 2: 629–630.

Bartel DP. (2009). MicroRNAs: target recognition and regulatory functions. Cell 136: 215–233.

Bartels CL, Tsongalis GJ. (2009). MicroRNAs: novel biomarkers for human cancer. Clin Chem 55: 623–631.

Boyd SD. (2008). Everything you wanted to know about small RNA but were afraid to ask. Lab Invest 88: 569–578.
Bracken CP, Gregory PA, Kolesnikoff N, Bert AG, Wang J, Shannon MF et al. (2008). A double-negative feedback loop between ZEB1-SIP1 and the microRNA-200 family regulates epithelial-mesenchymal transition. Cancer Res 68: 7846–7854.
Burk U, Schubert J, Wellner U, Schmalhofer O, Vincan E, Spaderna S et al. (2008). A reciprocal repression between ZEB1 and members of the miR-200 family promotes EMT and invasion in cancer cells. EMBO Rep 9: 582–589.
Coolen MW, Statham AL, Gardiner-Garden M, Clark SJ. (2007). Genomic profiling of CpG methylation and allelic specificity using quantitative high-throughput mass spectrometry: critical evaluation and improvements. Nucleic Acids Res 35: e119.
Corcoran DL, Pandit KV, Gordon B, Bhattacharjee A, Kaminski N, Benos PV. (2009). Features of mammalian microRNA promoters emerge from polymerase II chromatin immunoprecipitation data. PLoS One 4: e5279.
Cullen BR. (2004). Transcription and processing of human microRNA precursors. Mol Cell 16: 861–865.
Dykxhoorn DM, Wu Y, Xie H, Yu F, Lal A, Petrocca F et al. (2009). miR-200 enhances mouse breast cancer cell colonization to form distant metastases. PLoS One 4: e7181.
Foekens JA, Sieuwerts AM, Smid M, Look MP, de Weerd V, Boersma AW et al. (2008). Four miRNAs associated with aggressiveness of lymph node-negative, estrogen receptor-positive human breast cancer. Proc Natl Acad Sci U S A 105: 13021–13026.
Fujita S, Iba H. (2008). Putative promoter regions of miRNA genes involved in evolutionarily conserved regulatory systems among vertebrates. Bioinformatics 24: 303–308.
Gregory PA, Bert AG, Paterson EL, Barry SC, Tsykin A, Farshid G et al. (2008). The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. Nat Cell Biol 10: 593–601.
Han L, Witmer PD, Casey E, Valle D, Sukumar S. (2007). DNA methylation regulates MicroRNA expression. Cancer Biol Ther 6: 1284–1288.
Hiroki E, Akahira J, Suzuki F, Nagase S, Ito K, Suzuki T et al. (2010). Changes in microRNA expression levels correlate with clinicopathological features and prognoses in endometrial serous adenocarcinomas. Cancer Sci 101: 241–249.

Hu R, Dawood S, Holmes MD, Collins LC, Schnitt SJ, Cole K et al. (2011). Androgen receptor expression and breast cancer survival in postmenopausal women. Clin Cancer Res 17: 1867–1874.
Iorio MV, Ferracin M, Liu CG, Veronese A, Spizzo R, Sabbioni S et al. (2005). MicroRNA gene expression deregulation in human breast cancer. Cancer Res 65: 7065–7070.
Klinge CM. (2009). Estrogen regulation of microRNA expression.
Curr Genomics 10: 169–183.
Klug M, Rehli M. (2006). Functional analysis of promoter CpG methylation using a CpG-free luciferase reporter vector. Epigenetics 1: 127–130.
Korpal M, Kang Y. (2008). The emerging role of miR-200 family of microRNAs in epithelial-mesenchymal transition and cancer metastasis. RNA Biol 5: 115–119.
Lee JW, Park YA, Choi JJ, Lee YY, Kim CJ, Choi C et al. (2011). The expression of the miRNA-200 family in endometrial endometrioid carcinoma. Gynecol Oncol 120: 56–62.
Lehmann U, Hasemeier B, Christgen M, Muller M, Romermann D, Langer F et al. (2008). Epigenetic inactivation of microRNA gene hsa-mir-9-1 in human breast cancer. J Pathol 214: 17–24.Li A, Omura N, Hong SM, Vincent A, Walter K, Griffith M et al. (2010). Pancreatic cancers epigenetically silence SIP1 and hypo- methylate and overexpress miR-200a/200b in association with elevated circulating miR-200a and miR-200b levels. Cancer Res 70: 5226–5237.
Linhart C, Halperin Y, Shamir R. (2008). Transcription factor and microRNA motif discovery: the Amadeus platform and a compendium of metazoan target sets. Genome Res 18: 1180–1189.
Livak KJ, Schmittgen TD. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 25: 402–408.
Lodygin D, Tarasov V, Epanchintsev A, Berking C, Knyazeva T, Korner H et al. (2008). Inactivation of miR-34a by aberrant CpG methylation in multiple types of cancer. Cell Cycle 7: 2591–2600.
Lowery AJ, Miller N, Devaney A, McNeill RE, Davoren PA, Lemetre C et al. (2009). MicroRNA signatures predict oestrogen receptor, progesterone receptor and HER2/neu receptor status in breast cancer. Breast Cancer Res 11: R27.
Lujambio A, Ropero S, Ballestar E, Fraga MF, Cerrato C, Setien F et al. (2007). Genetic unmasking of an epigenetically silenced microRNA in human cancer cells. Cancer Res 67: 1424–1429.

Ma L, Teruya-Feldstein J, Weinberg RA. (2007). Tumour invasion and metastasis initiated by microRNA-10b in breast cancer. Nature 449: 682–688.
Maillot G, Lacroix-Triki M, Pierredon S, Gratadou L, Schmidt S, Benes V et al. (2009). Widespread estrogen-dependent repression of microRNAs involved in breast tumor cell growth. Cancer Res 69: 8332–8340.
Marson A, Levine SS, Cole MF, Frampton GM, Brambrink T, Johnstone S et al. (2008). Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. Cell 134: 521–533.
Mattie MD, Benz CC, Bowers J, Sensinger K, Wong L, Scott GK et al. (2006). Optimized high-throughput microRNA expression profiling provides novel biomarker assessment of clinical prostate and breast cancer biopsies. Mol Cancer 5: 24.
Mirnezami AH, Pickard K, Zhang L, Primrose JN, Packham G. (2009). MicroRNAs: key players in carcinogenesis and novel therapeutic targets. Eur J Surg Oncol 35: 339–347.
Monteys AM, Spengler RM, Wan J, Tecedor L, Lennox KA, Xing Y et al. (2010). Structure and activity of putative intronic miRNA promoters. Rna 16: 495–505.
Ozsolak F, Poling LL, Wang Z, Liu H, Liu XS, Roeder RG et al. (2008). Chromatin structure analyses identify miRNA promoters. Genes Dev 22: 3172–3183.
Schmidt D, Schwalie PC, Ross-Innes CS, Hurtado A, Brown GD, Carroll JS et al. (2010). A CTCF-independent role for cohesin in tissue-specific transcription. Genome Res 20: 578–588.
Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H et al. (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A 98: 10869–10874.
Visone R, Croce CM. (2009). MiRNAs and cancer. Am J Pathol 174: 1131–1138.
Waltering KK, Porkka KP, Jalava SE, Urbanucci A, Kohonen PJ, Latonen LM et al. (2011). Androgen regulation of micro-RNAs in prostate cancer. Prostate 71: 604–614.
Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C et al. (2008). Alternative isoform regulation in human tissue transcrip- tomes. Nature 456: 470–476.
Wang X, Xuan Z, Zhao X, Li Y, Zhang MQ. (2009). High-resolution human core-promoter prediction with CoreBoost_HM. Genome Res 19: 266–275.
Wiklund ED, Bramsen JB, Hulf T, Dyrskjot L, Ramanathan R, Hansen TB et al. (2011). Coordinated epigenetic repression of the miR-200 family and miR-205 in invasive bladder cancer. Int J Cancer 128: 1327–1334.

Wojdacz TK, Dobrovic A. (2007). Methylation-sensitive high resolu- tion melting (MS-HRM): a new approach for sensitive and high- throughput assessment of methylation. Nucleic Acids Res 35: e41.
Xu G, Wu J, Zhou L, Chen B, Sun Z, Zhao F et al. (2010). Characterization of the small RNA transcriptomes of androgen dependent and independent prostate cancer cell line by deep sequencing. PLoS One 5: e15519.
Yan HL, Xue G, Mei Q, Wang YZ, Ding FX, Liu MF et al. (2009). Repression of the miR-17-92 cluster by p53 has an important function in hypoxia-induced apoptosis. EMBO J 28: 2719–2732.
Yan LX, Huang XF, Shao Q, Huang MY, Deng L, Wu QL et al. (2008). MicroRNA miR-21 overexpression in human breast cancer is associated with advanced clinical stage, lymph node metastasis and patient poor prognosis. RNA 14: 2348–2360.
Zhou X, Ruan J, Wang G, Zhang W. (2007). Characterization and identification of microRNA core promoters in four model species. PLoS Comput Biol 3: e37.

Supplemental Data

Contains supplemental materials and methods, five tables and three figures.

Supplemental Materials and Methods

Generation of plasmid constructs

All promoter reporter constructs were cloned into pGL3-Basic (Promega) unless otherwise specified. A list of primers used is provided in Supplementary Tables 2 and 3. The miR-200b minigene was cloned by removing an XbaI digest fragment from the P2 construct and replacing the gap with a 2925bp PCR amplified genomic DNA fragment -2643bp to +282bp relative to the 5' stem loop of miR-200b. The minigene with the minimal promoter deleted was made by designing primers that would PCR amplify outward from the minimal promoter. This 7.5kb PCR fragment was then allowed to self-ligate. For the in vitro methylation plasmids, the BamHI/HindIII fragment of the P2 construct was subcloned into a CpG-free luciferase reporter construct pCpG-basic while P1 was PCR amplified and inserted into the HindIII site of the pCpG-basic plasmid (a gift from Klug et al). PCR was performed using KapaHiFi polymerase (Kapa Biosystems). All constructs were confirmed by sequencing.

Transfections and reporter assays

For luciferase assays, either MCF7 or MDA-MB-231cells were grown to 60-70% confluence in 24-well plates prior to transfections. Cells were co-transfected with 400ng of promoter construct and 10ng of RL-TK plasmid (Promega) as a transfection control. Transfected cells were incubated at 37°C and harvested for luciferase assays after 48hrs. Luciferase assays were performed using the Dual-Glo luciferase Assay kit (Promega) as recommended by the manufacturer. Luminescence was detected on a Beckman DTX plate reader. Firefly luciferase levels were then normalized to Renilla luciferase levels and expressed relative to pGL3-Basic levels (RLU). Statistical analysis was performed using unpaired two-tailed t-test.

For minigene experiments, either MCF7 or MDA-MB-231 cells were grown to 60-70% confluence in 6-well plates, and transfected with 1□g of DNA. Cells were left to incubate at 37°C for 72hrs. In all transfections, a 3□l:1□g ratio of Fugene (Roche) tranfection reagent to DNA was used.

Bisulfite modification and methylation sensitive high resolution melt analysis.

2μg of DNA extracted from cell lines was subjected to bisulfite modification using the MethylEasy Xceed kit (Human Genetic Signatures) according to the manufacturer's instructions. PCR amplification and MS-HRM was performed in duplicate on the RotorGeneTM Q (Qiagen). Primers were designed according to the principles outlined (Wojdacz and Dobrovic, 2007) to control for PCR bias and are shown in Supplementary Table 4 and 5. PCR was performed in a total volume of 20µl containing: 1x Buffer, 4mM MgCl, 200µM of each dNTP, 200nM of each primer, 5µM Syto9 dye (Invitrogen), 0.5U HotStarTaq polymerase (Qiagen) and 2ng of bisulphite modified template. The amplification conditions consisted of 15min at 95°C, followed by 50 cycles of 20sec at 95°C, 30sec at the appropriate annealing temperature (Tm) and 30sec at 72°C, followed by one cycle of 1min 30sec at 72°C and an MS-HRM step from 70°C to 90°C rising by 0.1°C/sec. Bisulfite treated CpGenomeTM Universal Methylated DNA (Chemicon, Millipore) and DNA from the appropriate cell line were used as positive/methylated and negative/unmethylated controls, respectively. WGA DNA made with the GenomiPhi kit (Amersham) was used for miR335 and miR663. To create a range of methylated dilutions, these controls were mixed in 25, 50 and 75% methylated to unmethylated template ratios. These standards were included in the analysis of each region.

Patient samples

Human breast tumours and matching lymph node metastases were collected from 56 patients, as approved by local Human Ethics committees, who underwent surgical resection and did not undergo preoperative radiochemotherapy at Princess Alexandra Hospital between 1988 and 2000. All patients were female aged from 30-94 years old, with a median age of 56 years. The tissues were formalin fixed paraffin embedded (FFPE) with routine histological examination using hematoxylin and eosin (H&E) confirming all as invasive Grade III ductal adenocarcinomas. Tumour-rich tissue in each biopsy was distinguished from surrounding normal tissue in H&E-stained sections by a qualified breast pathologist. 4µM sections were cut from the TMA blocks, transferred on to glass slides, deparaffinised and immunostained using anti-ER (SP1), anti-PR, (SP2), anti-HER2/neu (4B5) antibodies (pre-diluted; Ventana Medical Systems) the anti-AR antibody (Biocare Medical, 1:50), and counterstained with 3,3'-Diaminobenzidine (DAB) and hematoxylin. Staining was performed with the BenchMark® automated slide stainer (Ventana) using the iVIEW® DAB detection kit with additional Avidin and Biotin Blockers according to the

manufacturer's instructions. Analysis of strained sections was performed by a qualified breast pathologist. The presence of tumour tissue was first confirmed by examining the counterstain, and then all tumour sections were given a score for ER, PR, HER2 expression.

Supplementary figure legends

Table 1: List of miRNAs implicated in breast cancer. Sequences -5kb to +1kb of the 5' stemloop were assessed for CGIs. Strong CGIs are denoted in green, weak CGIs in yellow and non-CGIs in red. Expression levels in normal and cancer tissue are indicated in green for low expression and red for high expression. Known methylated miRNAs are labelled as methylated.

Table 2-5: Lists of primers used and their genomic coordinates on UCSC genome browser. The human genome NCBI36/ Hg18 March 2006 build was used.

Figure 1: Methylation levels determined by MS-HRM and miRNA expression in breast cancer cell lines

A: DNA methylation and miR-335 expression levels in a panel of 9 breast cancer cell lines.
B: DNA methylation and miR-10b expression levels in a panel of 9 breast cancer cell lines.
C: DNA methylation and miR-9 expression levels in a panel of 9 breast cancer cell lines.
D: DNA methylation and miR-663 expression levels in a panel of 9 breast cancer cell lines.
E: DNA methylation and miR-124 expression levels in a panel of 9 breast cancer cell lines.
F: DNA methylation and miR-373 expression levels in a panel of 9 breast cancer cell lines.
Top Panels: Black bars represent the percentage methylation. Bottom Panels: miRNA expression was determined by qPCR. Expression is shown relative to RNU6B and bars represent the mean + standard deviation of two independent experiments.
G: MS-HRM profile of the miR-210 promoter
H: MS-HRM profile of the miR-320a promoter

Figure 2A-E: miR-200b P1 and P2 MS-HRM profiles in MDA-MB-436, MDA-MB-468, T47D, ZR-75-1, MCF7, MDA-MB-157, MCF10a, HS-578-T and MDA-MB-231 cells. Blue, violet, red, orange, khaki lines represent 100%, 75%, 50%, 25% and 0% methylation standards respectively. The melt profile of the cell line is in black.

Figure 3. Hypermethylation is associated with loss of miR-200b in primary tumours.
(A)Top: Average P1 methylation in primary tumours. Bottom: miR200b expression.
(B)Top: Average P2 methylation in primary tumours. Bottom: Relative miR-200b levels in tumours.

Supplementary Table 1

miRNA ID	-5K CpG	-4k CpG	-3k CpG	-2k CpG	-1k CpG	+1k CpG	Normal Exp	BC exp	BC methylation Status
hsa-mir-124-1							up	down	methylated
hsa-mir-124-3							up	down	methylated
hsa-mir-191							down	up	
hsa-mir-126							up	down	
hsa-mir-196b							unknown	unknown	
hsa-let-7i							down	up	
hsa-mir-210							down	up	
hsa-mir-148a							up	down	methylated
hsa-mir-663							up	down	methylated
hsa-mir-9-1							up	down	methylated
hsa-mir-127							up	down	
hsa-mir-320a							up	down	
hsa-mir-124-2							up	down	methylated
hsa-mir-203							down	up	
hsa-mir-125b-1							up	down	
hsa-mir-136							down	up	
hsa-mir-301b							down	up	
hsa-mir-202							down	up	
hsa-mir-125a							up	down	
hsa-mir-365-1							up	down	
hsa-mir-7-3							down	up	
hsa-mir-152							up	down	methylated
hsa-mir-320b-2							up	down	
hsa-mir-200a							up	down	
hsa-mir-200b							up	down	
hsa-mir-429							up	down	
hsa-mir-24-2							unknown	unknown	
hsa-mir-27a							down	up	
hsa-mir-181d							down	up	
hsa-mir-17							up	down	
hsa-mir-18a							up	down	
hsa-mir-19a							up	down	
hsa-mir-19b-1							up	down	
hsa-mir-20a							up	down	
hsa-mir-196a-1							unknown	unknown	
hsa-mir-34a							up	down	methylated
hsa-mir-10b							down	up	
hsa-let-7a-3							up	down	

hsa-mir-141							up	down	
hsa-mir-200c							up	down	
hsa-mir-27b							up	down	
hsa-mir-365-2							up	down	
hsa-mir-24-1							unknown	unknown	
hsa-mir-373							down	up	
hsa-mir-196a-2							unknown	unknown	methylated
hsa-mir-497							up	down	
hsa-mir-320b-1							up	down	
hsa-mir-335							up	down	
hsa-mir-15b							unknown	unknown	
miRNA ID	-5K CpG	-4k CpG	-3k CpG	-2k CpG	-1k CpG	+1k CpG	Normal Exp	BC exp	BC BC methylation Status Status
hsa-mir-16-2							unknown	unknown	
hsa-mir-22							unknown	unknown	
hsa-mir-125b-2							up	down	
hsa-mir-451							down	up	
hsa-mir-21							unknown	unknown	
hsa-mir-143							up	down	
hsa-mir-145							up	down	
hsa-mir-489							up	down	
hsa-let-7a-1							up	down	
hsa-let-7f-2							up	down	
hsa-let-7a-2							up	down	
hsa-mir-101-1							up	down	
hsa-mir-204							up	down	
hsa-mir-205							up	down	
hsa-mir-206							up	down	
hsa-mir-30a							up	down	
hsa-mir-31							up	down	
hsa-mir-320c-1							up	down	
hsa-mir-320c-2							up	down	
hsa-mir-320d-1							up	down	
hsa-mir-320d-2							up	down	
hsa-mir-146b							unknown	unknown	
hsa-mir-16-1							unknown	unknown	
hsa-mir-146a							unknown	unknown	
hsa-mir-181a-1							down	up	
hsa-mir-181b-1							down	up	
hsa-mir-181b-2							down	up	
hsa-mir-222							down	up	
hsa-mir-221							down	up	
hsa-mir-29b-1							down	up	
hsa-mir-29b-2							down	up	
hsa-mir-29c							down	up	
hsa-let-7f-1							down	up	
hsa-mir-98							down	up	
hsa-mir-122							down	up	
hsa-mir-128-1							down	up	
hsa-mir-128-2							down	up	
hsa-mir-155							down	up	

hsa-mir-7-1							down	up	
hsa-mir-7-2							down	up	
hsa-mir-516a-1							down	up	
hsa-mir-516a-2							down	up	
hsa-mir-520c							down	up	

Supplementary Table 2: Primers for cloning miRNA promoters

Assay name	Primer sequences 5'→3'*	Position (UCSC Genome Browser, March 2006)	Amplicon size (bp)
miR124-1 prom1A Hind3 F	CCCAAGCTTGGGAACAGCGACGTCTTCCAAAG	chr8:9798613-9799100	488
miR124-1 prom1A Kpn1 R	GGGGTACCCCACTGCAGCAGGCGAGTTC		
miR124-1 prom1B Hind3 F	CCCAAGCTTGGGTGCAGCTCCAGACAATGAAA	chr8:9798752-9799100	308
miR124-1 prom2A Hind3 F	CCCAAGCTTGGGCACCAGCACACGTCATTCTC	chr8:9801030-9801870	841
miR124-1 prom2A Mlu1 R	CGACGCGTCGCACTTCTGCGCCTCTAATC		
miR124-1 prom2B R	AGTCAGGTGTCGATTTGACG	chr8:9801573-980187	339
miR124-2 prom1A Kpn1 F	GGGGTACCCCCTTTAACATTCCTTCC	chr8:65444604-65445773	1170
miR124-2 prom1A Hind3 R	CCCAAGCTTGGGCTCGTTGCCAGAAGTTGTTG		
miR124-2 prom1B Hind3 R	CCCAAGCTTGGGAGGCAGCTGTTTCCTCAGAA	chr8:65444604-65445031	428
miR124-2 prom1C Kpn1 F	GGGGTACCCCGCGCGGAGCTAGGCTGAG	chr8:65445362-65445773	412
miR124-2 prom2A Kpn1 F	GGGGTACCCCGGACTGCACAGAAGGACCAT	chr8:65452806-65453362	557
miR124-2 prom2A Hind3 R	CCCAAGCTTGGGAGGGTGAGGGAAGTTGACCT		
miR124-3 prom1A Kpn1 F	GGGGTACCCCAAGGGTCAAGAGGTGGTGAG	chr20:61276620-61277438	819
miR124-3 prom1A Hind3 R	CCCAAGCTTGGGCCTCTTGGAGTCCTGAGTAGC		
miR124-3 prom1B Hind3 R	CCCAAGCTTGGGCAACTTCGCTACGGGTCAG	chr20:61276620-61277438	292
miR124-3 prom1C Kpn1 F	GGGGTACCCCGTGTTTCCTCCCCTCCT	chr20:61276849-61277174	326
miR124-3 prom 1C Hind3 R	CCCAAGCTTGGGCTGGGGTCTGTGCTTC		
miR124-3 prom2A Kpn1 F	GGGGTACCCCGTTTCCTTGGGTCTCCGTGT	chr20:61279113-61279546	434
miR124-3 prom2A Hind3 R	CCCAAGCTTGGGAAGGGAGCCAGGCAAGT		
miR124-3 prom3A Kpn1 F	GGGGTACCCCATCCGTCTTCGCGATTCC	chr20:61280027-61280333	307
miR124-3 prom3A Hind3 R	CCCAAGCTTGGGAATCAAGGTCCGCTGTGAAC		
miR24-2 prom1A Hind3 F	CCCAAGCTTGGGACAACTCTCTGCCCACCTTG	chr19:13814191-13814881	691
miR24-2 prom1A Kpn1 R	GGGGTACCCCGGGGGACAGGACAGAGAAAC		
miR24-2 prom2 Hind3 F	CCCAAGCTTGGGACAGGAAGCAAATCCCATCC	chr19:13808435-13809052	617
miR24-2 prom2 Kpn1 R	GGGGTACCCCGGCCTGTATCTTGGAGCTTG		
miR125a prom1A Kpn1 F	GGGGTACCCCTCTGCGTCAGGCTTCTCTG	chr19:56884218-56884704	487
miR125a prom1A Hind3 R	CCCAAGCTTGGGCTCAGGGACCCAGGAGTG		

miR663 prom1A Hind3 F	CCCAAGCTTGGGCGTGATTCTCGTCCATCCTC	chr20:26137015-26137328	248
miR663 prom1A Kpn1 R	GGGGTACCCCGCTCGTCGCCTACTGTGG		
miR663 prom2A Hind3 F	CCCAAGCTTGGGCACTCAACCGCCTCGAAC	chr20:26138069-26138411	565
miR663 prom2A Kpn1 R	GGGGTACCCCTTACGTGGCAGCACTCTTG		
miR373 prom1A Kpn1 F	GGGGTACCCCAACCTGCGGAGAAGATACC	chr19:58982822-58983365	543
miR373 prom1A Hind3 R	CCCAAGCTTGGGCCGTGTTAGCCAGGATGG		
miR373 prom1B Hind3 R	CCCAAGCTTGGGCATCCGTTGATATGGGC	chr19:58982822-58983047	226
miR373 prom1C Kpn1 F	GGGGTACCCCGCCCATATCAACGGATGC	chr19:58983030-58983365	336
miR22 prom1A Kpn1 F	GGGGTACCCCTCGCCTGCTCTTTAGGACTC	chr17:1565993-1566592	600
miR22 prom1A Hind3 R	CCCAAGCTTGGGTCCTCGTAGCTCCTGACACAC		
miR22 prom1B Hind3 R	CCCAAGCTTGGGCTCCAGGCTCCGATCAGC	chr17:1566277-1566592	316
miR22 prom1C Kpn1 F	GGGGTACCCCAGCTGATCGGAGCCTGGAG	chr17:1565993-1566295	303
miR22 prom2A Kpn1 F	GGGGTACCCCTTGGACTGAGTGTCAGCA	chr17:1564534-1564954	421
miR22 prom2A Hind3 R	CCCAAGCTTGGGCGTTTGAAAGAAGGTCAGC		
miR9-3 prom1A Kpn1 F	GGGGTACCCCGCGTCCCTAAACCTTGTCAC	chr15:87711583-87712265	683
miR9-3 prom1A Hind3 R	CCCAAGCTTGGGAGAAACGGGCCTCCCTTAG		
miR9-3 prom1B Kpn1 F	GGGGTACCCCGATCCCTGGACTGACGT	chr15:87711848-87712047	200
miR9-3 prom1B Hind3 R	CCCAAGCTTGGGCTCCCCTGCTCCCCGTTC		
miR9-3 prom1C Kpn1 F	GGGGTACCCGGGGAGCAGGGGAGAAAT	chr15:87712034-87712265	230
miR193b prom1A Kpn1 F	GGGGTACCCCTTGTCTGGGCTGCGATTG	chr16:14304079-14304666	588
miR193b prom1A Hind3 R	CCCAAGCTTGGGAGAAACCAGAAACGCCACTC		
miR193b prom2A Kpn1 F	GGGGTACCCCGCCCCCTGTTTGAAGCAC	chr16:14303238-14303914	677
miR193b prom2A Hind3 R	CCCAAGCTTGGGAAGAAGGGGAGCACTCA		
miR193b prom2B Hind3 R	CCCAAGCTTGGGCTGCACCCGGCCTCCAC	chr16:14303607-14303914	397
miR10b prom1A Kpn1 F	GGGGTACCCCAAAGGGAGGCTGAATTGCTC	chr2:176720135-176720974	840
miR10b prom1A Bgl2 R	GAAGATCTTCTAACAGAAGTGAGCGCCTTG		
miR10b prom1B Kpn1 F	GGGGTACCCCGAGTGGGCGGATAGAAGAAA	chr2:176720708-176720974	267
miR10b prom2A Kpn1 F	GGGGTACCCCGAAGGTAAATGCGCGACTTC	chr2:176722816-176723177	362
miR10b prom2A Hind3 R	CCCAAGCTTGGGCAGGTAGCCTAATGGGCTTG		

miR301b prom1A Kpn1 F	GGGGTACCCCAAAAGGCGGGCAAAGTTC	chr22:20336232-20337076	845
miR301b prom1A Hind3 R	CCCAAGCTTGGGACCCCAGACTAGGGGGTTG		
miR301b prom1B Hind3 R	CCCAAGCTTGGGAGGGTTAGACGGAAGAGG	chr22:20336232-20336583	352
miR9-1 prom1A Kpn1 F	TTTTGGTACCAACGCCTTTCTGAGTTG	chr1:154656653-154657412	760
miR9-1 prom1A Bgl2 R	AAAAAGATCTTGTATCCTCTGGTGCTGGTC		
miR9-1 prom1B Bgl2 R	AAAAGGTACCTCAACTCCACTCGTGTCC	chr1:154656653-154656970	318
miR9-1 prom2A Kpn1 F	AAAAGGTACCAGGCAGCAAGAGGCTGAG	chr1:154657818-154658415	598
miR9-1 prom2A Bgl2 R	AAAAAGATCTGCTCTAGGGGTGGGAAAG		
miR9-1 prom2B Kpn1 F	AAAAGGTACCTCGGCGGAGGCTAAGAG	chr1:154657818-154658118	301
miR9-1 prom2C Bgl2 R	AAAAAGATCTCTTAGCCTCCGCCGAG	chr1:154658101-154658416	316
miR17 prom1A Kpn1 F	TTTTGGTACCGTCGAGTCCCAGGGAGAG	chr13:90797558-90798148	591
miR17 prom1A Bgl2 R	TTTTAGATCTGCCTGCGCTTTACTACGAC		
miR17 prom1B Kpn1 F	TTTTGGTACCTAATGAGGGAGTGGGGCTTG	chr13:90797886-90798148	263
miR17 prom1C Kpn1 F	TTTTGGTACCGCGCGCAGAGCTTGTTA	chr13:90797642-90797968	327
miR17 prom1C Bgl2 R	TTTTAGATCTGCACCTCGAAGGACCATGTG		
miR17 prom2A Kpn1 F	AAAAGGTACCAAGTGGAAGCCAGAAGAGGAG	chr13:90800420-90800746	327
miR17 prom2A Bgl2 R	AAAAAGATCTGCATAATCCCTAATGGGGAAG		
miR210 prom1A Kpn1 F	AAAAGGTACCAAACCAGGCAGAGCCAGAG	chr11:558382-559119	738
miR210 prom1A Bgl2 R	AAAAAGATCTCTCGGCCCAACTTCAG		
miR210 prom1B Kpn1 F	AAAAGGTACCAGACGTGCAGAAAAGAACG	chr11:558382-558631	250
miR210 prom1C Bgl2 R	AAAAAGATCTCGTTCTTTTCTGCACGTCTG	chr11:558611-559119	509
miR210 prom1D Kpn1 F	AAAAGGTACCGAGCCCGGGCCATACCAC	chr11:558528-558777	250
miR210 prom1D Bgl2 R	AAAAAGATCTACCGCAACGCAGCCAGTG		
miR210 prom 1E Bgl2 R	AAAAAGATCTCTCGGCCCAACTTCAG	chr11:558382-558777	396
miR320a prom1A F	AGAGGGGTAGGCTTGAGAGG	chr8:22158500-22159077	578
miR320a prom1A Bgl2 R	AAAAAGATCTAGCGCCGCCTGATAAATAC		
miR320a prom1B Kpn1 F	CCCCGGTACCCAGGTGAGAGCCTTTG	chr8:22158500-22158752	253

miR320a prom1C Bgl2 R	AAAAAGATCTCAAAGGCTCTCACCTGGGT	chr8:22158743-22159087	345
miR335 prom1A Kpn1 F	TTTTGGTACCCCTCTCACGGTTCAGTACCC	chr7:129918665-129919256	592
miR335 prom1A Bgl2 R	TTTTAGATCTGGCAGCTACAGCCACTCC		
miR335 prom1B Blg2 F	TTTTAGATCTAACTCATCAGGGGAGGGTTT	chr7:129918879-129919256	378
miR335 prom1C Bgl2 R	TTTTAGATCTGAGGTGCCGGGGTGT	chr7:129918879-129919192	314
miR335 prom1D Bgl2 R	TTTTAGATCTCGACTTTTAGAGCCCACCC	chr7:129918879-129919142	264

Supplementary Table 3: Primers for miRNA-200b promoter experiments

Assay name	Primer sequences 5'→3'*	Position (UCSC Genome Browser, March 2006)	Amplicon size (bp)
miR200b P1 Kpn1 F	TTTTGGTACCGCAGAGGTGGAGAGGCGAGAGT	chr1:1087797-1088137	341
miR200b P1 Hind3 R	TTTTAAGCTTGGGGCCTCGGGAGGGAAGAG		
miR200b P2 Kpn1 F	TTTTGGTACCTTTTACAGCCCGGATCACTG	chr1:1089316-1090354	1039
miR200b P2 Hind3 R	TTTTAAGCTTCGCTTTCTTGTCAACCGTC		
P1+P2 Hind3 F	TTTTAAGCTTGGGAGGCAGAGGTGGAGAG	chr1:1087791-1090354	2564
P1+P2 Hind3 R	TTTTAAGCTTCGCTTTCTTGTCAACCGTC		
-2447/-1993bp Hind3 F	TTTTAAGCTTACGCAGAGGGAAGAACCTG	chr1:1089901-1090354	454
-2447/-1993bp Hind3 R	TTTTAAGCTTCGCTTTCTTGTCAACCGTC		
-2228/-1993bp Hind3 F	TTTTAAGCTTAGCCTGTGCAGGTGGGAC	chr1:1090120-1090354	235
-2228/-1993bp Hind3 R	TTTTAAGCTTCGCTTTCTTGTCAACCGTC		
-2228/-2037bp Hind3 F	TTTTAAGCTTAGCCTGTGCAGGTGGGAC	chr1:1090120-1090310	191
-2228/-2037bp Hind3 R	TTTTAAGCTTGTGCCGGTTGAGGTGTTG		
-3032/-2447bp Kpn1 F	TTTTGGTACCTTTTACAGCCCGGATCACTG	chr1:1089316-1089900	585
-3032/-2447bp Kpn1 R	TTTTGGTACCCAGGTTCTTCCCTCTGCGT		
miR-200b minigene Kpn1 F	TTTTGGTACCTTTTACAGCCCGGATCACTG	chr1:1089315-1092628	3314
miR-200b minigene Xba1 R	TTTTTCTAGAATTCCGGGGTCTCTGAGATG		
miR-200b minigene Promoter KO F	GACGGTTGACAAGAAAGCG		
miR-200b minigene Promoter KO R	GTCCCACCTGCACAGGCT		
P1 in vitro methylation F	TTTTAAGCTTGGGAGGCAGAGGTGGAGAG		
P1 in vitro methylation R	TTTTAAGCTTGGGGCCTCGGGAGGGAAGAG		
5'PCR Walk primer A	TTTTACAGCCCGGATCACTG		

5'PCR Walk primer B	ACGCAGAGGGAAGAACCTG
5'PCR Walk primer C	AGCCTGTGCAGGTGGGAC
5'PCR Walk primer D	GACGGTTGACAAGAAAGCG
5'PCR Walk primer R1	GCCTTATGCAGTTGCTCTCC
5'PCR Walk primer R2	TCTTCCAGCGGATAGAATGG

Supplementary Table 4: Primers for miR-200b DNA methylation experiments

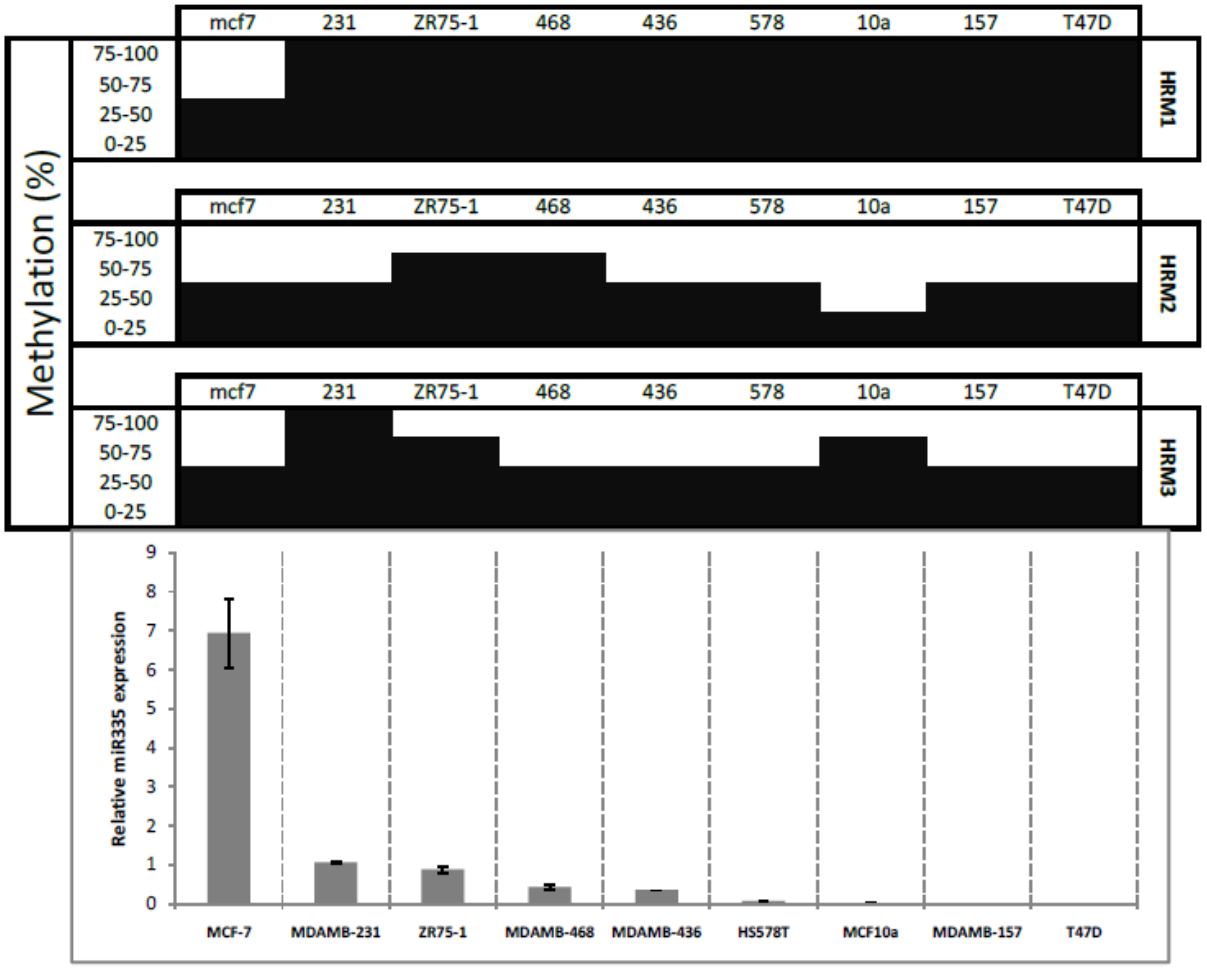
Assay name	Primer sequences 5'→3'*	Position (UCSC Genome Browser, March 2006)	Amplicon size (bp)	# CpG assessed	Tm (˚C)
P1 MS HRM F	GGCGGGGAGTATTGTTTTTTG	chr:1088073-1088347	275	16	64
P1 MS HRM R	CCGATCCACGAAACTAAATACTCTACC				
P2 MS HRM 1 F	CGTATTTTTGGATTTTTGGAGGAGT	chr:1089737-1090055	319	27	60
P2 MS HRM 1 R	AACCTACCCGACGAACTTAATCAATA				
P2 MS HRM 2 F	GTCGGGCGTTTTTATTTTATTTTAGTT	chr:1090095-1090224	130	8	62
P2 MS HRM 2 R	CGCCCCAACAAAAAATTCTCTA				
P1 Sequenom F	AGGAAGAGAGTTATGGGAGTTTAGGGGAT ATATT	chr1:1087998-1088221	224	17	60
P1 Sequenom R	CAGTAATACGACTCACTATAGGGAGAAGGC TATTCAAACCTACACAAATAAA				
P2 Sequenom F	AGGAAGAGAGGGGGAGGGTTGGATTTTAT AT	chr1:1089990-1090222	233	17	60
P2 Sequenom R	CAGTAATACGACTCACTATAGGGAGAAGGC TCCCCAACAAAAAATTCTCTA				

Supplementary Table 5: Primers for all other miRNA methylation experiments

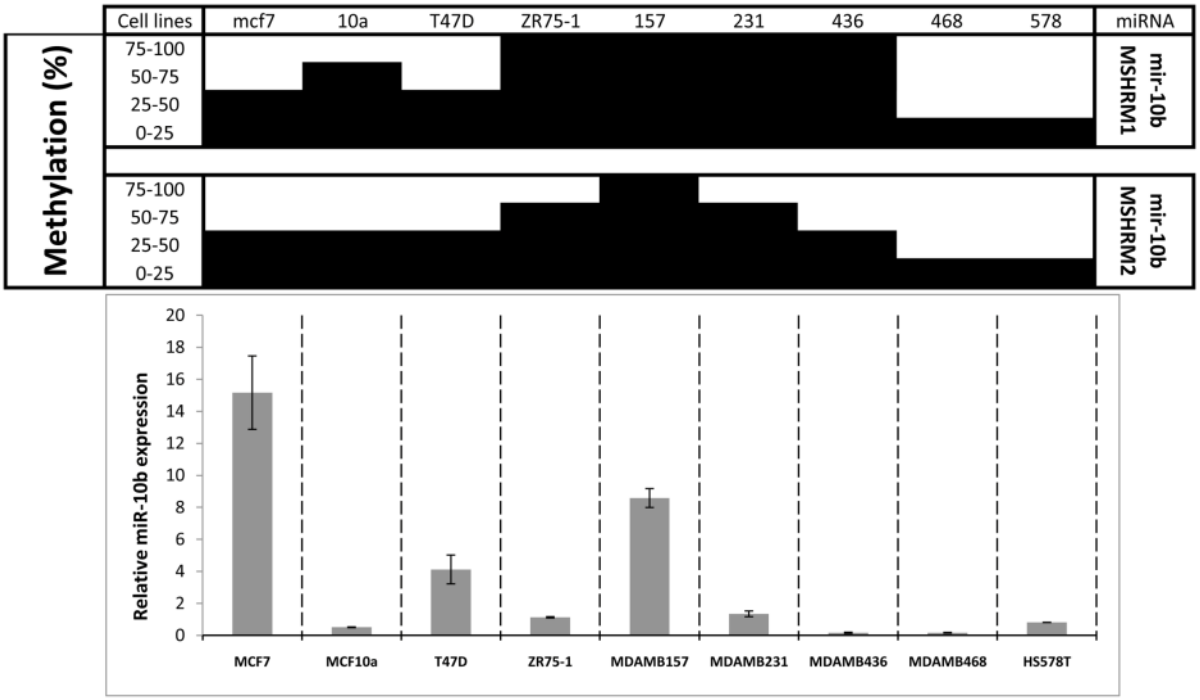
Assay name	Primer sequences 5'→3'*	Position (UCSC Genome Browser, March 2006)	Amplicon size (bp)	# CpG assessed	Tm (˚C)
mir124-3 MSHRM F	TTTAGGATTGAGAGGTTTGGAGGAT	chr20:61278843-61278948	106	10	60
mir124-3 MSHRM R	CCRCACTCRAACCCTAACC				
mir10b MSHRM-1 F	GTTTAAGGGTGYGTGGAGTG	chr2:176720352-176720468	118	6	61
mir10b MSHRM-1 R	CCCRAAAACTACRCAAAATTC				
mir10b MSHRM-2 F	TTTGTAGAATYGAATTTGTGTGGTATT	chr2:176723306-176723425	120	5	60
mir10b MSHRM-2 R	CCATATCRCACTTTAATCTCTAACTATTC				
mir9-3 MSHRM-1 F	GTTTTYGGTTTTTTTGTGGAGA	chr15:87711542-87711645	104	5	60
mir9-3 MSHRM-1 R	ACCCACTACRAACCATCAAAA				
mir9-3 MSHRM-2 F	GGGYGTTYGAGGTTTTTTAAG	chr1: 87711933-87712057	126	15	59
mir9-3 MSHRM-2 R	CRACRCATTTCTCCCCTAC				
mir663 MSHRM F	TTTTGygGTGTTTTTGGA	chr20: 26138005-26138144	139	14	60
mir663 MSHRM R	CRACRAACTCCCTCAAAAC				

mir373 MSHRM F	AACCACAAACTCTTTAATTCCTACAAAAA	chr19: 58982977-58983089	113	7	60
mir373 MSHRM R	TTATTTTGATGTTTAAGTGGAAAGTGTTG				
mir124-1 MSHRM F	TGGTTGGGTYGGTTGAAT	chr8:9800144-9800270	127	7	60
mir124-1 MSHRM R	CTCCTACRCRTCCCTTCTC				
mir124-2 MSHRM F	GTAGGGAGAYGATTAGGTTTGTTT	chr8:65444916-65445035	120	7	60
mir124-2 MSHRM R	CCCRAAACAACTATTTCCTCAA				
mir9-1 MSHRM F	GCGGTAGAGTTAATTAGAGGATGGTT	chr1:154658114-154658265	152	8	60
mir9-1 MSHRM R	CCGAACCTAAACGAACAAAATAAAA				
mir210 MSHRM F	GGTCGGTTATTGGTTGAGGGATT	chr11:558379-558512	134	12	60
mir210 MSHRM R	CGACTCTCGACCCAACTTCAA				
mir320a MSHRM F	TCGCGCGTTATAATTTTATTTGT	chr8:22158504-22158660	157	16	58
mir320a MSHRM R	CCGCCTAATAAATACTATAACCCAA				
mir335 MSHRM-1 F	TTAYGTAAATAAAGGGGGTTTTGTTT	chr7:129918319-129918430	112	8	60
mir335 MSHRM-1 R	CAAACRCCTAACCACAACAAC				
mir335 MSHRM-2 F	GGTTTTAATTTATTAGGGGAGGGTTTT	chr7:129918873-129919015	143	5	60
mir335 MSHRM-2 R	AAATACCCCTCTAACAATAACTCCTTTAAA				
mir335 MSHRM-3 F	GYGGTGTAGTTTAGGATTTTAAGATTTAG	chr7:129920121-129920245	125	6	60
mir335 MSHRM-3 R	CTCACTCRCRCTTATTACTAAAATCTA				

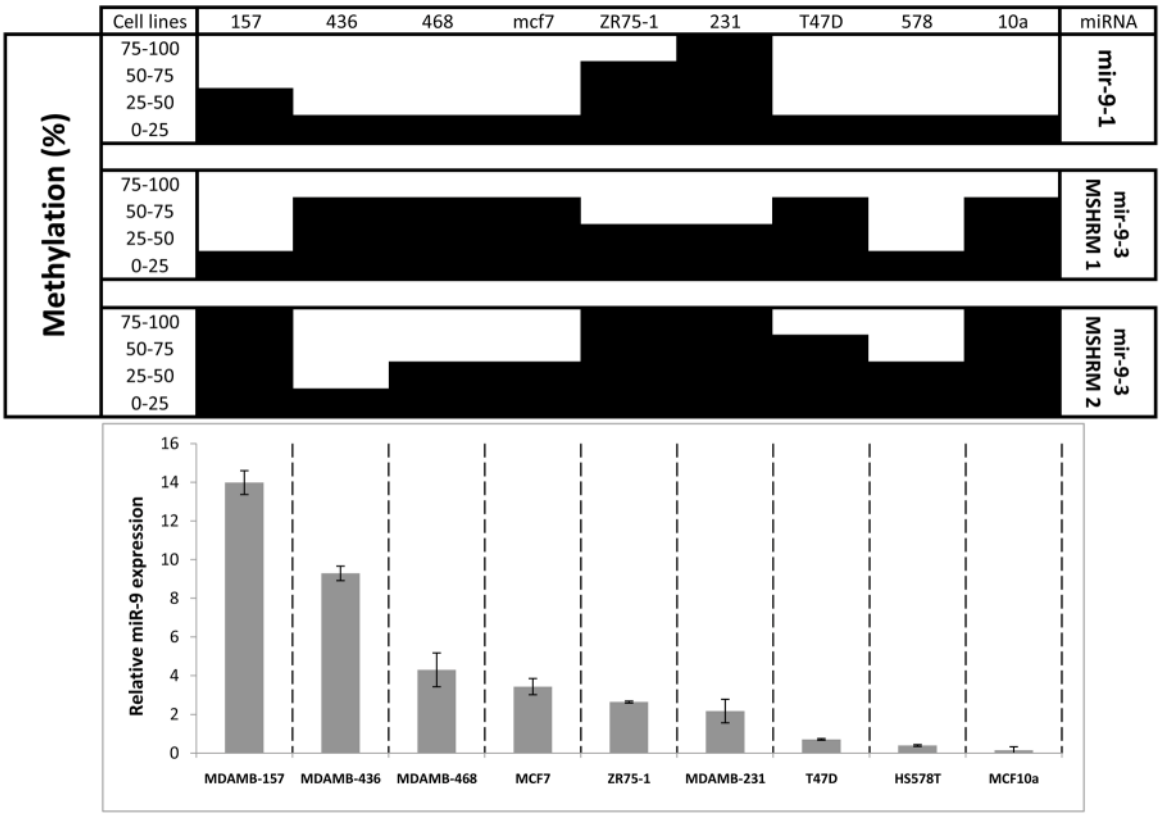
Supplementary Figure 1A



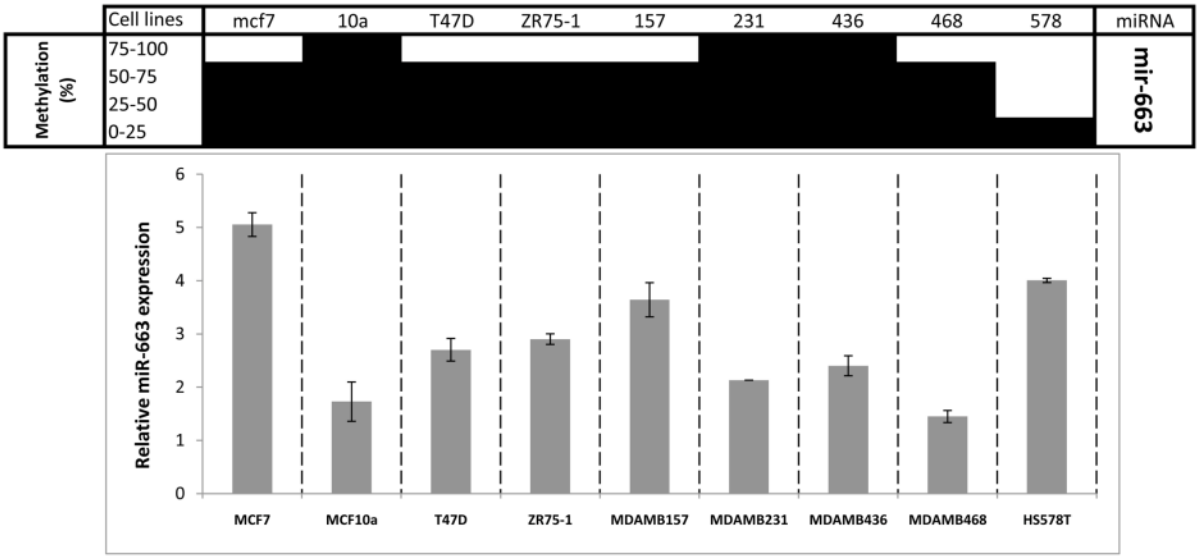
Supplementary Figure 1B



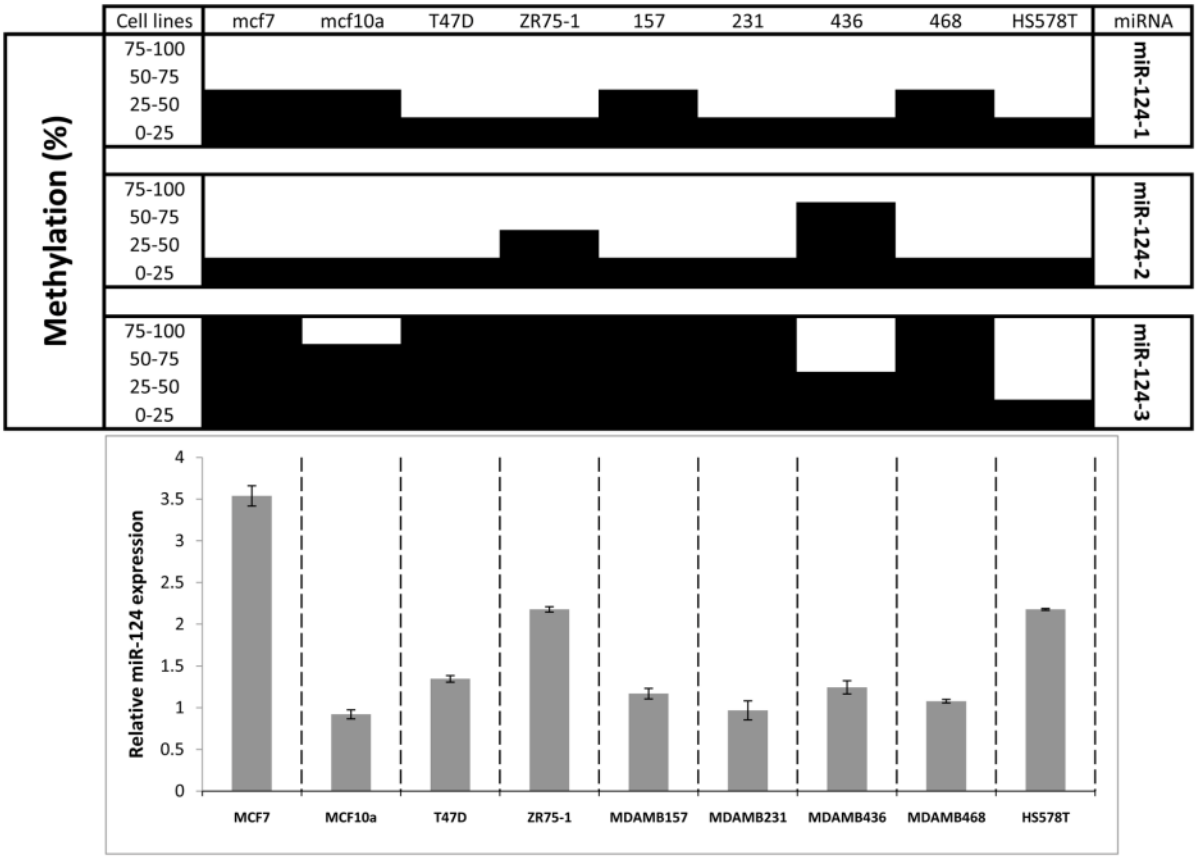
Supplementary Figure 1C



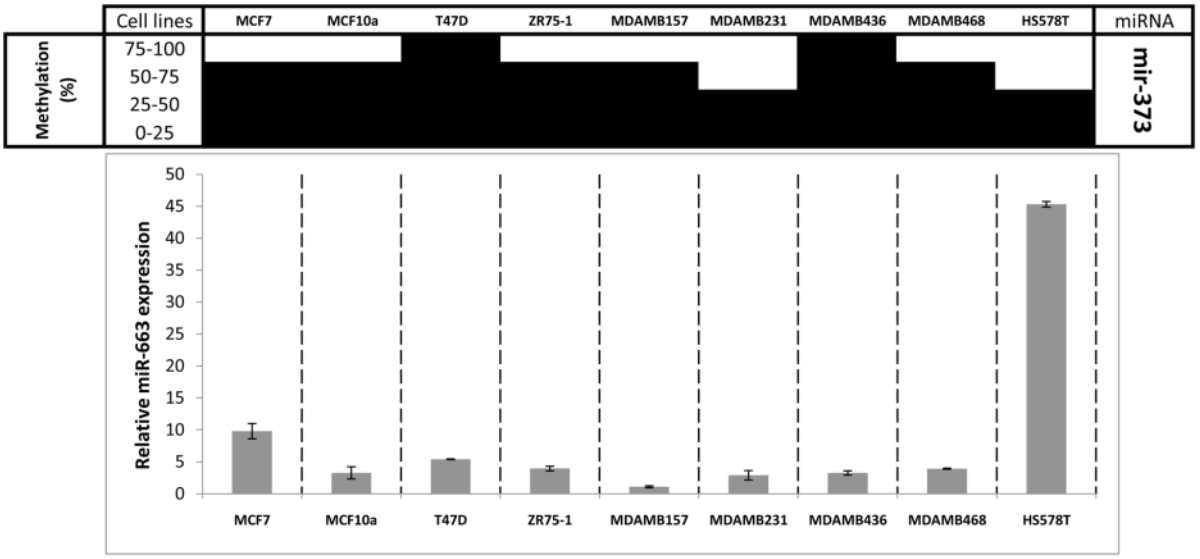
Supplementary Figure 1D



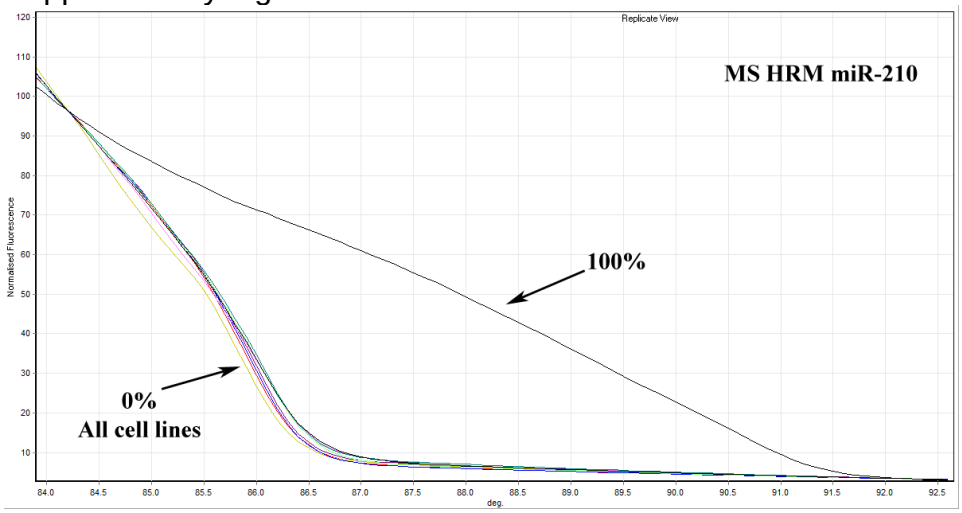
Supplementary Figure 1E



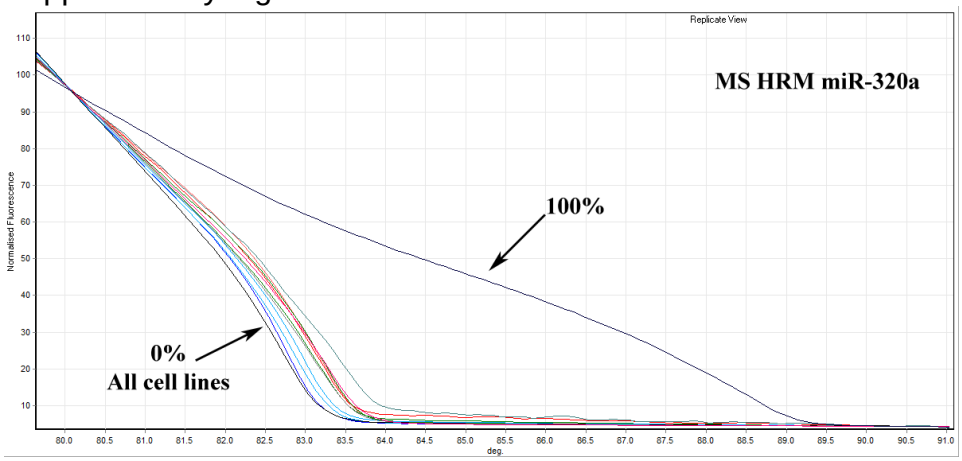
Supplementary Figure 1F



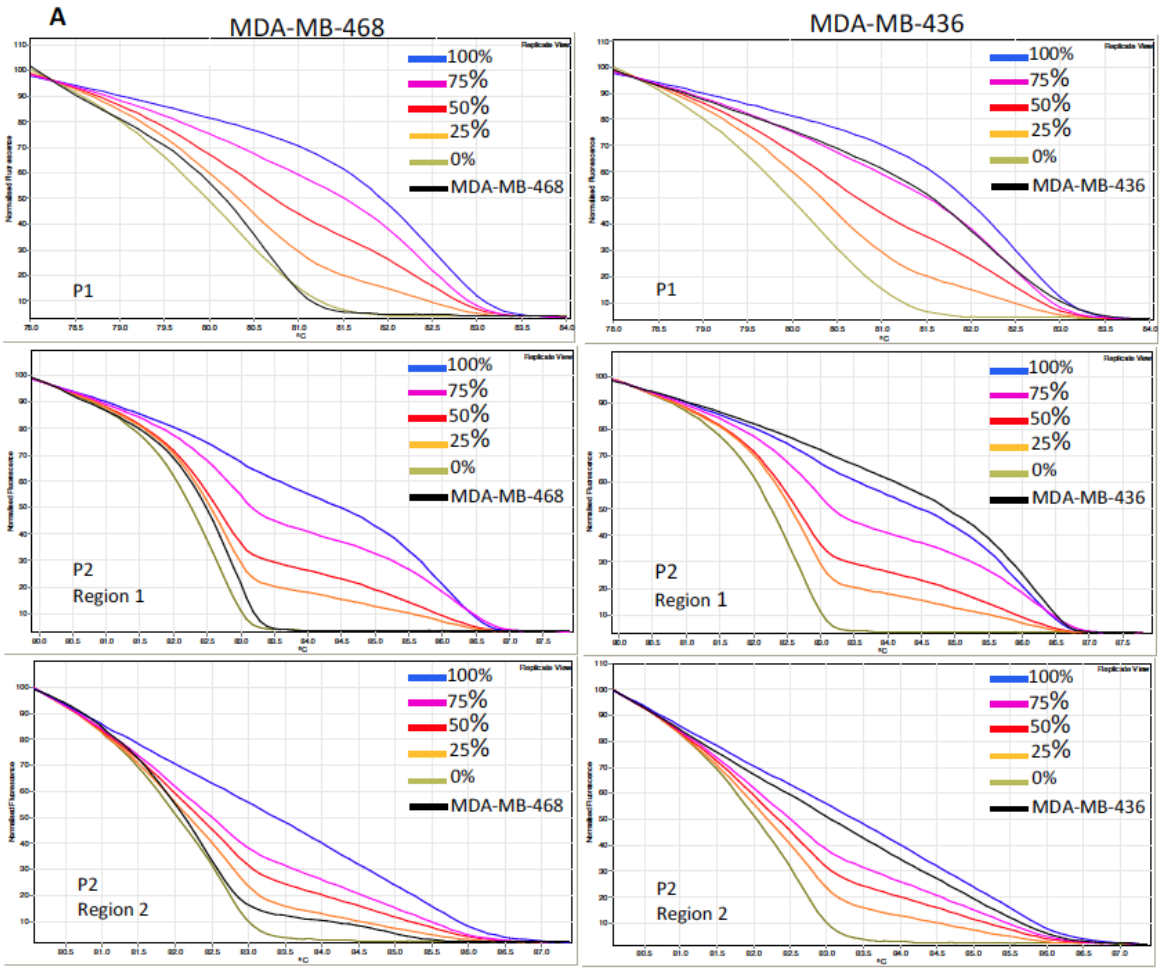
Supplementary Figure 1G



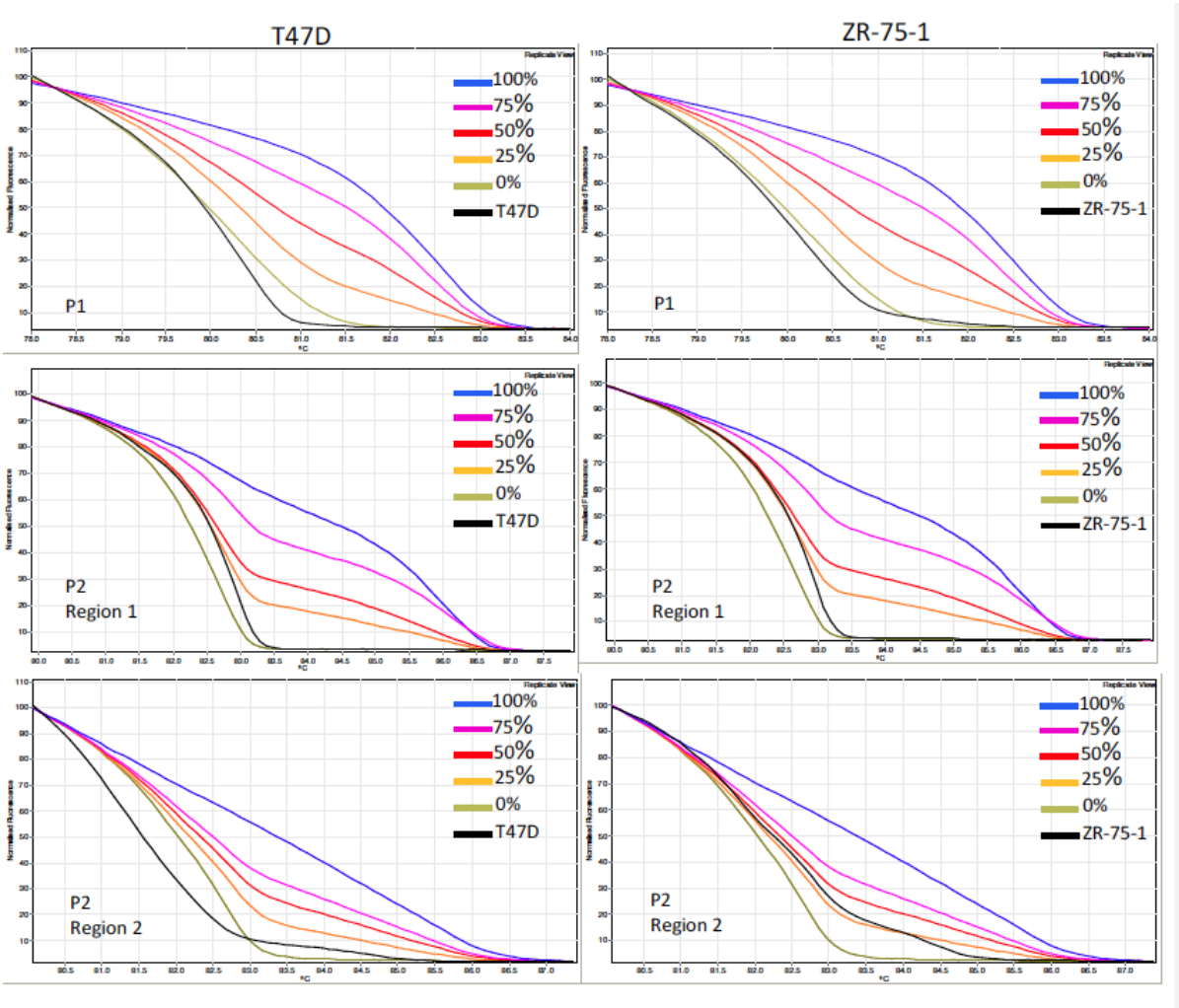
Supplementary Figure 1H



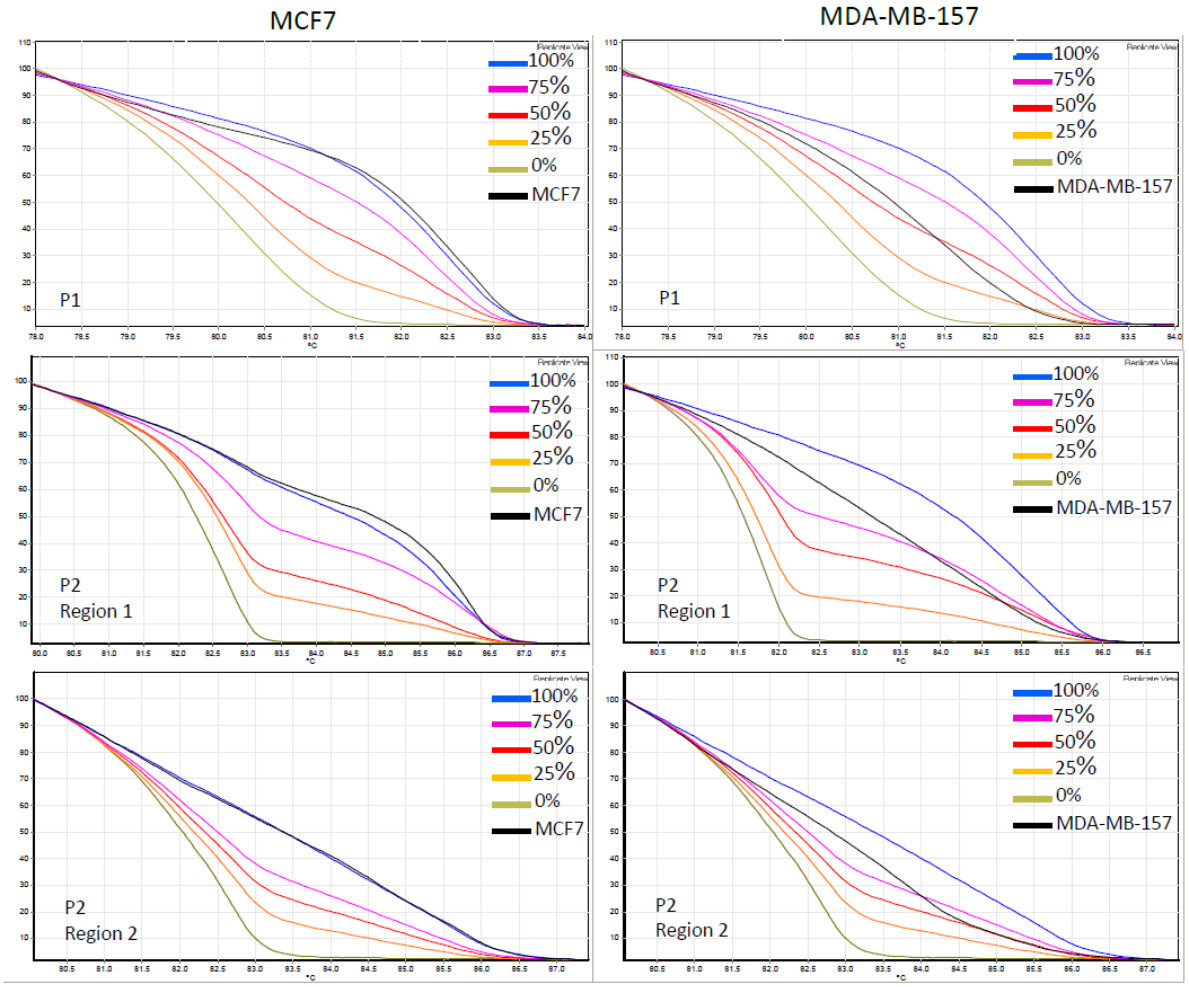
Supplementary Figure 2A



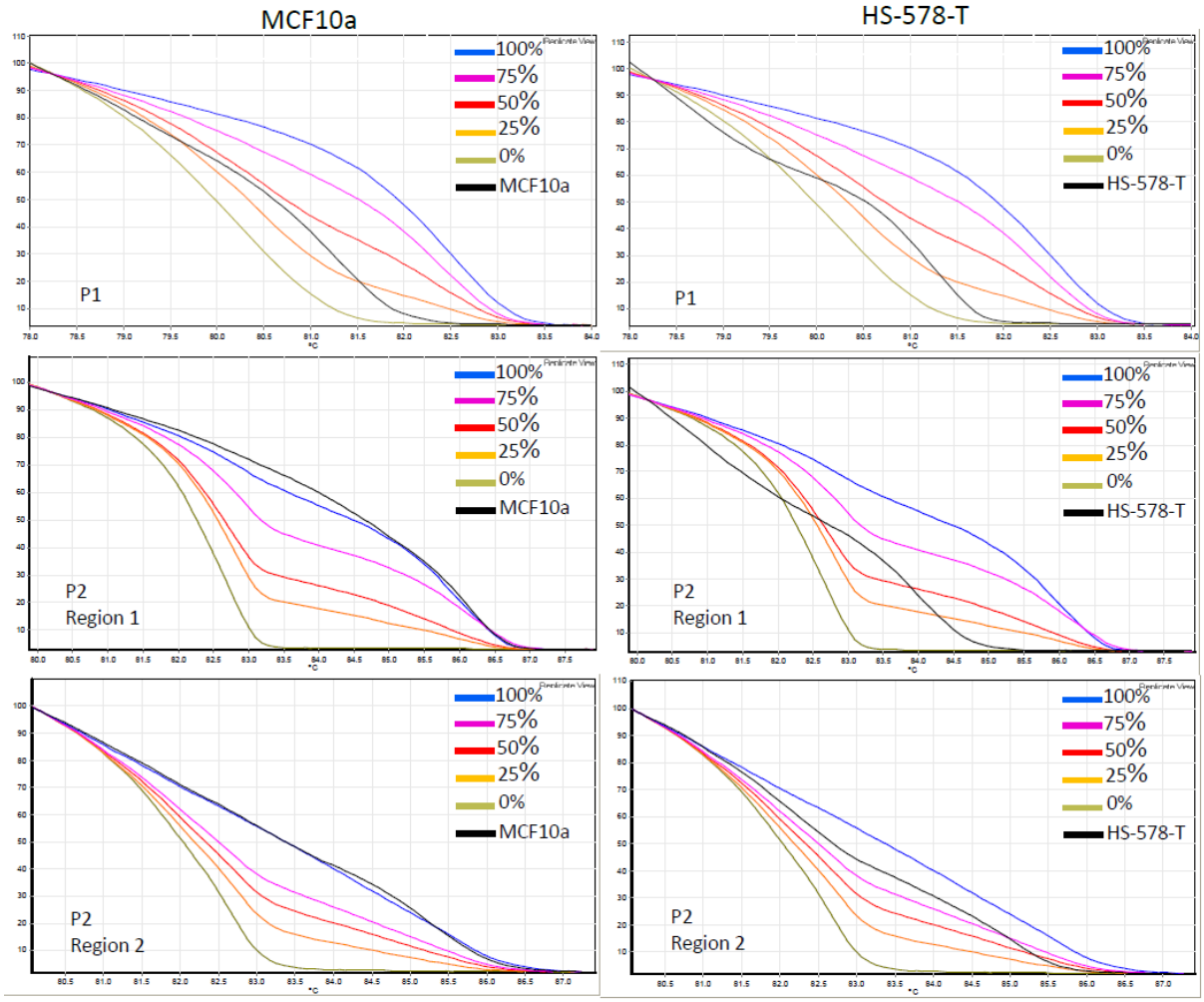
Supplementary Figure 2B



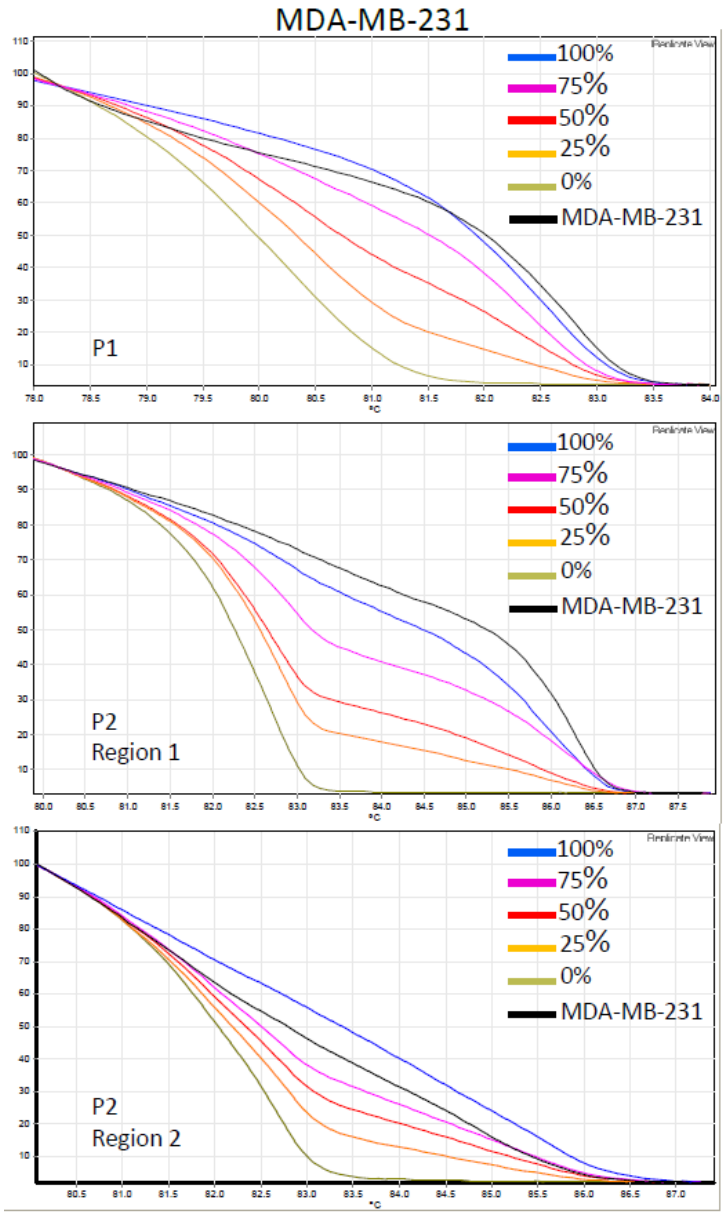
Supplementary Figure 2C



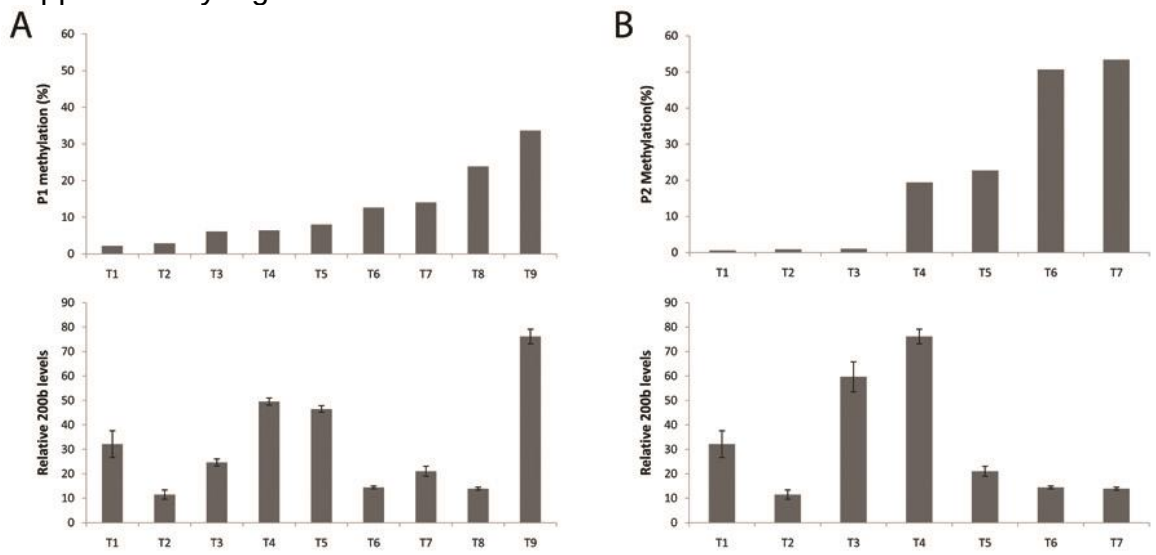
Supplementary Figure 2D



Supplementary Figure 2E



Supplementary Figure 3



Chapter 5

WADDELL, N., STEIN, S. R., WAGNER, S. A., BENNETT, I., DJOUGARIAN, A., MELANA, S., JAFFER, S., HOLLAND, J. F., POGO, B. G., GONDA, T. J., BROWN, M. A., LEO, P., SAUNDERS, N. A., MCMILLAN, N. A., COCCIARDI, S., VARGAS, A. C., LAKHANI, S. R., CHENEVIX-TRENCH, G., NEWMAN, B. & FRANCIS, G. D. 2012. Morphological and molecular analysis of a breast cancer cluster at the ABC Studio in Toowong. Pathology, 44, 469-72.

6.1.3. Introduction

The study cohort comprised women who worked at the ABC studios in Toowong who developed breast cancer between 1994 and 2007. Standard demographic details of the patients were extracted from the pathology record and clinical information was collated. Tissue microarrays were constructed from the FFPE blocks including both patient samples and control specimens using a semiautomated tissue microarrayer, Beecher Galileo. Material was not available for 1 patient; 8 patients had material suitable for inclusion in the TMA and 27 controls samples were included. 4 µm sections were cut from the TMA and immunohistochemistry was performed for ER, PR, HER2 (Ventana 4B5), CK 5/6, SMA, p63, HPV and SV40. In situ hybridization testing was performed for HER2 (Ventana Inform, Ventana Medical Systems) and EBV.

DNA was extracted using needle macro dissection and a lysis protocol. RNA (500 ng) was extracted from 2-10 sections/sample (mean number of sections was 4.6). The Illumina Whole Genome-DASL (cDNA-mediated Annealing, Selection, Extension, and Ligation) assay which measure approximately 24,000 transcripts was performed as per manufacturer's instructions. Unsupervised Hierarchical Clustering whereby transcripts which had a >2 fold change versus the mean were clustered. Pair wise analysis to determine differential expression was performed using the linear model and empirical Bayes method. Coded specimens of the patient samples and controls were used for MMTV analysis.

6.1.4. Results

Sir,

Cancer clusters involving small numbers of patients are relatively common, but only a few have been investigated. In the case of breast cancer, for some putative clusters no significant increase in the risk of developing cancer has been found,[1] and in only a minority of cancer clusters which are found to be statistically significant is a causative agent found. Two such studies have shown an increased risk in breast cancer in office workers exposed to strong magnetic fields.[2,3] However, for many cancer clusters, although a common aetiology was not proven, it could not be discounted, highlighting the difficulty in definitively associating a causal agent to a cancer cluster.[4,5]

Nevertheless, in 2004 when staff at the Australian Broadcasting Corporation (ABC) premises in Toowong, Queensland, raised concerns about the number of breast cancer

cases occurring among female employees at that site, investigations were initiated. In 2005, two studies attempted to evaluate the likelihood of the ABC cancer cluster and to identify whether personal breast cancer risk factors or workplace exposure to electromagnetic fields were possible causal agents, but no firm conclusions could be drawn.[6] When an additional patient was diagnosed with breast cancer in 2006, an independent panel was constituted for a more systematic inquiry. That review included 550 women employed full- or part-time during the period from 1 January 1994 to 30 June 2006. A total of 13 women were diagnosed with breast cancer, with a possible additional case, and 10 of these were diagnosed while still working at the Toowong studio.[7] The expected number based on the Queensland rate of breast cancer was 1.6 cases.[6] The likelihood of this occurring by chance was estimated at approximately one in a million and this represented a greater than 6-fold increased risk compared to the general female population. No specific cause of the cluster was identified, despite measurements of electromagnetic fields, ionising radiation, and chemical contamination of water and soil, as well as traditional breast cancer risk factors.[7]

In an attempt to elucidate the nature of this cluster, we instigated a review of the biology of the ABC patients' breast tumours. Breast cancer is a heterogeneous disease composed of several molecular pathology subtypes which may arise via different biological pathways. The working hypothesis was that a comparison to molecular subtypes may show that the tumours within the cancer cluster would have more similar morphological and/or molecular characteristics than those seen among tumours from a comparison group, and that this might reflect a common exposure to an unknown factor which might, in the long term, help to identify a novel breast cancer risk factor. This study also included an analysis of viral agents as there are some data to suggest a role for a retrovirus like mouse mammary tumour virus (MMTV) in breast cancer.[8,9]

Fourteen patients from the ABC breast cancer cluster were contacted and consent was obtained to study their tumours. All available samples were in the form of formalin fixed, paraffin embedded (FFPE) blocks and histological slides. For each ABC case, up to three controls, matched for year of diagnosis and age of diagnosis (+/– one year), were retrieved from a large database at the Princess Alexandra Hospital. Ethics approval for the project was obtained from the Princess Alexandra Hospital Human Research Ethics Committee, with access to patient records approved by the Director General of Queensland Health under legislative requirements. Slides were de-identified and reviewed independently by two pathologists (SRL, GF). A proforma was developed to aid systematic reporting and the

primary tumour characteristics were recorded for each sample. No slides were available for two ABC cases, no FFPE blocks were available for two cases and no material, even a pathology report, was available for one case (Table 1). Of the remaining 12 cases, one case had ductal carcinoma in situ (DCIS) only, and one tumour was too small for additional studies. Eight ABC cases had material suitable for inclusion in the tissue microarray (TMA). Twenty-nine control tumours were also included with two excluded (one was DCIS and another was too small for inclusion) (Table 1).

TMAs were constructed in duplicate from the case and control FFPE blocks. Four mm sections were cut from the TMA and immunohistochemistry was performed using standard protocols. TMA results were available on six cases with no tumour or loss of the cores in two cases. Scoring of the sections was based on the percentage of cells stained (0–100%) and intensity of staining (0, 1, 2, 3,) for IHC. Oestrogen receptor (ER) and progesterone receptor (PR) cut-points were 1% or more of positive nuclei for classification of ER positive.[10]

Sufficient material was available for MMTV-like agent analysis from seven ABC cases and 18 control samples for MMTV-like envelope (env) gene analysis by polymerase chain reaction (PCR).[8]

For the gene expression analysis, the sections were cut on the day of RNA extraction and macrodissected with a needle to enrich for areas with >50% tumour cells. RNA was extracted with a standard Tri reagent protocol. The Illumina Whole Genome-DASL (cDNA-mediated annealing, selection, extension, and ligation) assay which measures approximately 24 000 transcripts was performed according to the manufacturer's instructions (Illumina, USA). Failed samples (n=3) were removed from analysis due to low average signal intensity (<100). Data were quantile normalised and unsupervised hierarchical clustering whereby transcripts which had a

>2 fold change versus the mean were clustered. Pair-wise analysis to determine differential expression was performed using the linear model and empirical Bayes method with Benjamini and Hochberg multiple testing correction.[11] Single sample predictor was used to determine which sporadic molecular breast tumour subtype each sample resembled by comparing to the five centroid training set of 249 breast tumour samples from literature.[12]

The results showed that no statistically significant differences were identified between the ABC case and control groups for any of the clinicopathological variables assessed (Table 2). The majority of the case and control tumours were ductal carcinoma of no special type.

Tumour grade and size were similar between the case and control groups, as was the distribution of molecular subtypes as predicted by immunohistochemistry (IHC). No cases or controls showed staining with HPV or EBV in situ hybridisation, nor were they positive with IHC for SV40 or HPV. MMTV-like env gene was identified in 66.7% of case samples and 59.3% of control samples but this difference was not statistically significant (p=1). RNA was extracted from eight cases and 26 controls, and for three of these tumours (two cases and one control) two independent RNA extractions were performed from different areas of the tumour. Where possible the WG-DASL assay was performed in duplicate for each RNA sample. However, for three samples (one case and two controls) insufficient RNA was available and so these samples were assayed once only. An additional three samples (one case and two controls) failed quality control and were removed from analysis. In total, 68 samples were analysed successfully: eight cases (four in duplicate, two in quadruplicate, two singly) and 25 controls (22 in duplicate, one in quadruplicate and two singly). We found only 13 genes had a significant difference in expression between cases and controls ($p=<0.05$ with Benjamini and Hochberg multiple testing correction). The gene expression profiles of all 68 samples were clustered using unsupervised hierarchical clustering (Fig. 1). The proportion of molecular subtypes present among the cases and controls was determined (Table 3). The classification was discordant for six of 29 duplicate pairs, but for three of these six samples the classification was either luminal A or luminal B. We assigned the final classification using the sample with the strongest prediction. There were no significant differences in the frequency of specific subtypes in cases versus controls. Fifty percent of the cases (4/8) were classified as luminal A compared to 36% (9/25) controls. There were no luminal B or basal-like tumours within the small set of ABC cases on gene expression profiling. Table 4 shows the relationship between the gene expression and immunohistochemistry subtypes.

In conclusion, a clinicopathological review of the breast cancers showed no significant differences between tumour samples from the ABC cases and matched controls. The molecular subtyping showed that there were fewer luminal B and basal subtypes in the cases than controls, which does not support the hypothesis that BRCA1-carriers (which usually have the basal subtype) could account for the cluster.[7] Viral agents were investigated as possible aetiological factors for the ABC cluster because there are some data to suggest a role for MMTV-like agent in breast cancer.[8,9] However, there were no statistically significant differences in the prevalence of any virus between the ABC case and control groups.

There are no differences between the ABC cases and control tumours at the molecular level. Gene expression profiling is able to stratify tumours based upon phenotypic characteristics, but expression of only a few genes differed between the ABC cases and the controls, and unsupervised hierarchal clustering did not separate the tumours from the cases and controls. A supervised cluster analysis of the intrinsic genes that differentiate the breast tumour subtypes also showed that the ABC cases and controls did not cluster separately (data not shown). Together these findings indicate that the ABC cases and control tumours do not differ at a molecular level, and show that the ABC tumours are similar to the previously characterised sporadic molecular subtypes.

A high level of exposure to an exogenous agent in the environment might account for the cluster, but the environ- mental analyses performed previously found no evidence for such agents (associated reports are available at http://www. abc.net.au/corp/pubs/bci.htm). In 2009, a report was released on the risk of breast cancer among female employees of the ABC studios throughout Australia, including 5969 women employed between January 1994 and December 2005. This study reconfirmed the excess rate of breast cancer among ABC employees in Queensland, but no excess rate was observed among staff in the rest of Australia. A recent editorial[4] reviewed an updated analysis of the ABC cluster[13] which reported a much higher probability of the initial cluster than the one in a million originally reported, and concluded that the absence of an increased risk of breast cancer at other ABC sites in Australia, together with a failure to identify any causative agent and a revised probability of 1 in 25, means that the situation may have arisen by chance. Our results showing no shared pattern of morphological or molecular characteristics are consistent with such a conclusion.

In summary, the breast cancers from the ABC cluster appear to be morphologically and molecularly similar to the range of breast cancers seen in the general population based on pathological review, immunohistochemistry subtyping, presence of various viruses and gene expression profiling. The limitations of our analyses of the tissue samples include the size of the tumours and the availability only of FFPE tissue; no frozen material or blood samples were available for the cases from the time of diagnosis. With the development of new technology, such as next generation sequencing, the possibility exists for a more in-depth analysis of the samples in the future. At this point in time, it seems reasonable to conclude that the cancers among the women of the ABC cluster are not unique. However, it should be noted that we cannot rule out a carcinogenic influence.

Acknowledgement: The authors gratefully acknowledge participation of the women employed by the ABC Studio in Toowong who developed breast cancer.

Conflicts of interest and sources of funding: Financial support was provided by the National Breast Cancer Fund's Novel Concept Scheme. The authors state that there are no conflicts of interest to disclose.

Contact Dr G. Francis.
Email: pathcare@msn.com.au

Table 1: Summary of samples

	Slides available for review	FFPE material available	Included in TMA	RNA extracted
Case	12	12	8	8
Control	31	31	29	26
Total	43	43	37	34

Table 2: Clinicopathological characteristics of ABC case and control groups

	ABC Case Group		Control Group		p-value
Characteristic	n	%	n	%	
Age at Diagnosis					
50 years or less	11	79	26	84	
> 50 years	3	21	5	16	
Mean Age	42.79		44.35		
Tumour type					
Ductal NOS	11	79%	26	84%	
Ductal/Lobular	0	0%	1	3%	
Mucinous	1	7%	1	3%	
Micropapillary	0	0%	2	6%	
Unknown	1	7%	0	0%	
DCIS	1	7%	1	3%	
Histological Grade					0.34
GI	2	14%	8	26%	
GII	5	36%	10	32%	
GIII	4	29%	12	39%	
DCIS	1	7%	1	3%	

Unknown	2	14%	0	0%	
Tumour Size					0.09
<10mm	3	21%	4	13%	
10-20mm	7	50%	12	39%	
>20-50mm	1	7%	11	35%	
>50mm	1	7%	3	10%	
Unknown	1	7%	0	0%	
DCIS	1	7%	1	3%	
Hormone receptor status					1.0
ER+PR+	10	71%	19	61%	
ER+PR-	0	0%	1	3%	
ER-PR+	0	0%	0	0%	
ER-PR-	1	7%	10	32%	
Unknown	2	14%	0	0%	
DCIS	1	7%	1	3%	
IHC subtype					0.10
Luminal A	7	50%	19	61%	
Luminal B	3	21%	1	3%	
HER2	1	7%	5	16%	
Triple Negative	0	0%	5	16%	
DCIS only	1	7%	1	3%	
Unknown	2	14%	0	0%	
Lymph node status					0.78
Positive	6	43%	14	45%	
Isolated tumour cells	0	0	1	3%	
Negative	6	43%	15*	48%	
Unknown	2*	14%	1	3%	

*One was DCIS only

Table 3. Proportion of molecular subtypes within the ABC cases and controls determined by gene expression profiling

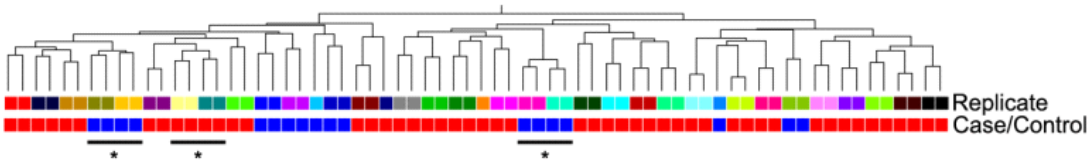
	Basal-like	**HER2**	**Luminal A**	**Luminal B**	**Normal-like**	**TOTAL**
Control	8 (32%)	1 (4%)	9 (36%)	4 (16%)	3 (12%)	25 (100%)
Case	0	1 (12.5%)	4 (50%)	0	3 (37.5%)	8 (100%)

Table 4. Overall correlation between gene expression profiling and IHC subtypes.

	IHC subtype			
Gene expression Subtype	Triple Negative	HER2	Luminal A	Luminal B
Basal-like	5	2	1	0
HER2	0	1	0	1
Luminal A	0	0	13	0
Luminal B	0	0	4	0

Normal-like	0	1	3	2

Figure 1. Unsupervised clustering of all 68 samples successfully arrayed from 8 cases and 25 controls. Genes which varied 2-fold from the mean in at least 2/68 samples were clustered. Replicate samples are coloured similarly and cluster together; quadruplicate samples are marked with a bar and an asterix. Cases are shown in blue, and controls in red.



1. Mulla ZD. A report of a breast cancer cluster among employees of an elementary school. Breast J 2005; 11: 162– 3.

2. Milham S Jr. Increased incidence of cancer in a cohort of office workers exposed to strong magnetic fields. Am J Ind Med 1996; 30: 702– 4.

3. Milham S. A cluster of male breast cancer in office workers. Am J Ind Med 2004; 46: 86– 7.

4. Stewart BW. The ABC breast cancer cluster: the bad news about a good outcome. Med J Aust 2010; 192: 629– 31.

5. Thun MJ, Sinks T. Understanding cancer clusters. CA Cancer J Clin 2004; 54: 273–80.

6. Armstrong B. Breast Cancer at the ABC Toowong Queensland: Final Report of the Independent Review and Scientific Investigation Panel. 2 June 2007 (cited May 2012). http://www.abc.net.au/corp/pubs/documents/ Breast_Cancer_Toowong_Final_Report.pdf

7. Stewart BW. ‘‘There will be no more!’’: the legacy of the Toowong breast cancer cluster. Med J Aust 2007; 187: 178– 80.

8. Wang Y, Holland JF, Bleiweiss IJ, et al. Detection of mammary tumor virus env gene-like sequences in human breast cancer. Cancer Res 1995; 55: 5173– 9.

9. Zur Hausen H. The search for infectious causes of human cancers: where and why. Virology 2009; 392: 1– 10.

10. Blows FM, Driver KE, Schmidt MK, et al. Subtyping of breast cancer by immunohistochemistry to investigate a relationship between subtype and short and long term survival: a collaborative analysis of data for 10,159 cases from 12 studies. PLoS Med 2010; 7: e1000279.

11. Li J, Smyth P, Flavin R, et al. Comparison of miRNA expression patterns using total RNA extracted from matched samples of formalin- fixed paraffin-embedded (FFPE) cells and snap frozen cells. BMC Bio- technol 2007; 7: 36.
12. Hu Z, Fan C, Oh DS, et al. The molecular portraits of breast tumors are conserved across microarray platforms. BMC Genomics 2006; 7: 96.
13. Sitas F, O'Connell DL, van Kemenade CH, Short MW, Zhao K. Breast cancer risk among female employees of the Australian Broadcasting Corporation in Australia. Med J Aust 2010; 192: 651– 4.

Chapter 6

PETERS, K. M., EDWARDS, S. L., NAIR, S. S., FRENCH, J. D., BAILEY, P. J., SALKIELD, K., STEIN, S., WAGNER, S., FRANCIS, G. D., CLARK, S. J. & BROWN, M. A. 2012. Androgen receptor expression predicts breast cancer survival: the role of genetic and epigenetic events. BMC Cancer, 12, 132.

6.1.5. Introduction

In triple negative breast carcinoma (TNBC), reduced androgen receptor (AR) expression has been associated with lymph node metastases and a worse overall survival (Luo et al., 2010, Thike et al., 2014), however the correlation of AR with clinicopathological features and survival is complex (McNamara et al., 2014). This study confirmed that loss of AR expression is associated with a worse overall survival.

6.1.6. Results

Abstract

Breast cancer outcome, including response to therapy, risk of metastasis and survival, is difficult to predict using currently available methods, highlighting the urgent need for more informative biomarkers. Androgen receptor (AR) has been implicated in breast carcinogenesis however its potential to be an informative biomarker has yet to be fully explored. In this study, AR protein levels were determined in a cohort of 73 Grade III invasive breast ductal adenocarcinomas. AR was expressed in 56% of tumours and expression was significantly inversely associated with 10 year survival (P=0.004). An investigation into the mechanisms responsible for the loss of AR expression revealed that hypermethylation of the AR promoter is associated with loss of AR expression in breast cancer cells but not in primary breast tumours. In AR negative breast tumours, mutation screening identified the same mutation (T105A) in the 5'UTR of two AR negative breast cancer patients but not reported in the normal human population. Reporter assay analysis of this mutation however found no evidence for a negative impact on AR 5'UTR activity. The role of miR-124 in regulating AR expression was also investigated, however no evidence for this was found. This study highlights the potential for AR expression to be an informative biomarker for breast cancer survival and sets the scene for a more comprehensive investigation of the molecular basis of this phenomenon.

Key words: Androgen receptor; prognostic biomarker; breast cancer; gene regulation; promoter methylation; regulatory mutation; miRNA.

Background

Breast cancer is a heterogeneous disease comprising tumour subtypes associated with variable clinical characteristics (Sorlie et al., 2001). Variables including tumour size, histological subtype and grade, lymph node status and the expression of estrogen receptor alpha (ERα), progesterone receptor (PR) and human epidermal growth factor receptor 2 (HER2) currently assist routine clinical management (Rakha et al., 2010b). However, these factors are limited in their ability to predict individual survival and response to therapy (Rakha et al., 2010b). This is particularly apparent for patients with advanced breast cancer, which is characterised by high histological grade and the presence of lymph node metastases, and has an aggressive clinical course and generally a poor prognosis (Rakha et al., 2010b). Identifying new prognostic biomarkers and the molecular mechanisms underlying breast cancer progression are paramount for improving the clinical management of these patients and developing improved therapeutic strategies.

Androgen receptor (AR) is a member of the nuclear receptor superfamily and is known to be involved in a complex network of signalling pathways that collectively regulate cell proliferation (Yeh et al., 2003, Liao and Dickson, 2002). Expressed in the normal human mammary gland, where it predominantly localises to the inner layer of epithelial cells lining acini and intralobular ducts (Li et al., 2010), the role of AR in normal mammary epithelial biology is unknown. AR has been implicated in breast tumourigenesis, however delineating its precise function has proven difficult with AR-mediated androgenic effects shown to both stimulate and inhibit growth of breast cancer cells (Doane et al., 2006, Birrell et al., 1995). The significance of AR in human breast cancer is further emphasized by the recent finding that it can be targeted in estrogen receptor negative breast tumours (Ni et al., 2011). Loss of AR expression is associated with early onset, high nuclear grade and negative ER, PR and HER2 expression status in breast tumours (Agoff et al., 2003, Gonzalez-Angulo et al., 2009). However, the mechanisms responsible for this loss of AR expression in breast carcinogenesis remain unclear.

The AR gene comprises 9 exons spanning 180.25 kilobases located on chromosome Xq12. Functional analyses have identified two independently regulated transcription initiation sites (TIS), AR-TIS I (-12/-11/-10) and AR-TIS II (-1/+1) (Fig. 1) (Faber et al., 1991). Transcriptional initiation from AR-TIS I is dependent on sequences located between

positions -17 and +45 and initiation from AR-TIS II facilitated by a palindromic homopurine repeat and SP1 binding to a GC-box (Faber et al., 1993, Chen et al., 1997). Additional putative cis-acting elements include HL (helix-loop-helix-like) motifs 1 and 2 (Takane and McPhaul, 1996) and a cAMP responsive element (Mizokami et al., 1994). Two CpG islands (CGI) are also located in the AR promoter and extend into Exon 1. Hypermethylation of these CGI have been shown to silence AR transcription in prostate cancer cells and primary tumours (Kinoshita et al., 2000). Genetic alterations in the promoter and 5'untranslated regions (UTR) of the AR gene have been also observed in prostate cancer cell lines, xenografts (Waltering et al., 2006) and in two prostate cancer patients (Crocitto et al., 1997, Cox et al., 2006). In breast cancer, the role of regulatory defects in the AR gene are yet to be fully elucidated.

In this study, we show that loss of AR expression is significantly associated with poor 10 year survival outcome in Grade III invasive breast ductal adenocarcinomas. We then evaluated potential regulatory mechanisms that may account for the loss of AR expression. For the first time we show that DNA hypermethylation in the AR promoter is associated with loss of AR expression in breast cancer cells, although this is not the case in our cohort of tumours from patients with Stage III breast cancer. We subsequently assessed whether somatic mutations in AR regulatory regions or miRNAs bioinformatically predicted to target the human AR 3'UTR might contribute to the observed changes in AR expression.

Results and Discussion

Low AR protein levels are associated with poor 10 year survival in patients with Stage III breast cancer.

To assess the prognostic value of AR expression in breast cancer patients, IHC analysis was performed in a cohort of 73 Grade III lymph node positive ductal adenocarcinomas from patients with Stage III disease. Patient and tumour characteristics are summarised in Table 1. The patients ranged in age from 30 to 94 years (mean, 54 years); with the majority of patients (97%) aged over 35 years. AR expression was detected in 56% (n=41) of primary breast tumours. Positive expression of ER, PR and HER2 was also observed in 55.5% (n=40), 40% (n=29) and 21.7% (n=15) of breast tumours, respectively. In AR-negative tumours, the majority (72%, n=23; 87%, n=27; 86.6%, n=26) were also ER, PR and HER2 negative, respectively. The authors acknowledge the potential limitations of TMA analysis, given the inherent heterogeneity of tumour samples, but there is high concordance between TMA cores and whole sections (Kyndi, M., et al., Tissue microarrays compared with whole sections and biochemical analyses. A subgroup analysis of DBCG 82 b&c. Acta Oncol, 2008. 47(4): p. 591-9.). In addition the impact was further minimized by analysing at least two cores from each tumour, in accordance with the correlation nomograms developed by Karlsson et al, 2009 (Karlsson et al., 2009). AR expression was a significant prognostic factor for overall patient survival (P=0.004) (Figure 2). The 10-year survival of patients with AR positive tumours was 52% versus 22% for patients with AR negative tumours. This finding is consistent with previous studies in a diversity of breast cancer patient populations wherein a significant association between AR expression and age at diagnosis, nuclear grade, recurrence-free survival was observed (Castellano et al., 2010, Bryan et al., 1984, Gonzalez-Angulo et al., 2009, Agoff et al., 2003, Park et al., 2011).

DNA methylation of the AR promoter is associated with low AR mRNA levels in breast cancer cell lines.

To investigate potential mechanisms responsible for the loss of AR expression, we evaluated the methylation status of the AR promoter region and AR expression levels in breast cancer cell lines. DNA methylation was determined by MS-HRM analysis of bisulfite treated DNA in three regions; Regions 1 and 2 correspond to the CpG island in the AR minimal promoter (Takane and McPhaul, 1996) and Region 3 corresponds to the CpG

island at the translational start site (Fig. 1). DNA methylation was also assessed in a further region, Region 4, by Sequenom MassARRAY (Fig. 1). Cell lines, MDAMB231, MCF7, MDAMB157, MDAMB468 and MDAMB436 all showed between 25-100% methylation in at least 2 of regions analysed (Fig. 3a). Notably, methylation of the AR promoter region was associated with the level of AR mRNA (Fig. 3 a and b). A similar association has been observed in prostate cancer, where treatment of prostate cancer cell lines that display AR hypermethylation with the demethylating agent 5-aza-2'-deoxycytidine induces the re-expression and function of AR (Jarrard et al., 1998).

AR promoter methylation is not associated with low AR protein levels in primary breast tumours.

To examine whether AR promoter methylation is also associated with loss of AR protein levels in breast tumours, the DNA methylation status of Region 4, which contains six CpG dinucleotides, was assessed by Sequenom MassARRAY. Sequenom was chosen for this analysis as it can analyse methylation at each individual CpG dinucleotide, reliably detecting methylation as low as 5%, in a high-throughput manner (Coolen et al., 2007). Primer design constraints meant that Region 4 of the AR promoter (Fig. 1) was selected for analysis, with DNA methylation of overlapping Regions 1 and 2 associated with AR mRNA levels in breast cancer cell lines. DNA methylation was observed in breast cancer patients at each of the six CpGs (Fig. 4). However, with the exception of CpG's 1-3, at which methylation in most tumours was greater than 30%, for the most part only low level methylation (< 30%) was observed. Furthermore, no significant association was observed in the average methylation between AR negative and AR positive primary breast tumours in our cohort at any of the six CpG dinucleotides examined.

There are several plausible explanations for the lack of association between promoter methylation and expression in primary breast tumours. There is the possibility that a region outside that examined shows expression associated methylation and indeed, there was an incomplete association between Region 4 methylation and AR mRNA levels in breast cancer cells. Additionally, another mechanism such as somatic mutation of these regions, or aberrant targeting by a miRNA, may be involved.

Identification of somatic mutations in the AR 5'UTR in AR negative breast tumours.

Somatic mutation of regulatory regions of the AR gene is another potential mechanism responsible for reduced AR expression in breast tumours. To address this possibility, we sequenced the 5' regulatory region of AR (-659 to +280) in breast cancer cell lines. Our results revealed no sequence variations in MDAMB157, MDAMB231, MDAMB436, MDAMB468, MCF7, T47D, ZR75-1, Hs578T and BT549 cells. We also sequenced the AR promoter region from -6 to +133 in 32 primary breast tumours. Amplification of the AR promoter region from -165 to -7, which corresponds to the homopurine repeat and GC box (Fig. 1), revealed a mutation (mRNA pos 105, T>A, Fig. 1), mapping to the AR 5'UTR, in two patients. This sequence variation does not correspond to any known SNPs (GRCh37 reference primary assembly, www.ncbi.nlm.nih.gov/sites/entrez?db=snp) or, to our knowledge, any previously reported AR variants.

To investigate the potential significance of the AR 5'UTR T105A variant, we performed bioinformatics analysis on the wild-type and variant sequence. In the wild-type sequence the T position is invariant in mammals and is a component of the binding site for RNApolII (based on ChIP-seq data) and the predicted and conserved binding site for several transcription factors, including RUNX1, En1 and Pax6. Based on the currently available ChIP-seq data however, there is currently no evidence that these transcription factors bind to this sequence in vivo. The substitution from T to A results in an abolishment of these predicted sites and the creation of predicted and conserved binding sites for NHLH1 (data not shown).

To experimentally address the effect of this variant on 5'UTR activity, we fused the 5'UTR upstream of the firefly cDNA and downstream of either the AR or the SV40 promoter, in pGL3-basic and pGL3-promoter vectors, respectively. The AR 5'UTR T105A sequence variant did not have a negative impact on SV40 or AR driven reporter activity in either MCF7 or T47D cells (Fig. 5). Instead, the AR 5'UTR T105A sequence variant actually increased AR reporter driven activity in T47D cells (P=0.0001) (Fig. 5).

Somatic mutation analysis of AR negative breast tumours identified the same alteration (chrX:66680703, mRNA pos 105, T>A) in the AR 5'UTR of two breast tumours. Examination of the functional importance of this sequence variation however revealed that there was no negative impact on the activity of the AR 5'UTR in MCF7 or T47D cells. This suggests that this mutation is unlikely to negatively affect the regulation of AR expression.

However, particularly given the AR 5'UTR T105A sequence variant significantly increased AR reporter driven activity in T47D cells, more complex studies, such as analysing the consequence of this mutation in the context of the entire AR gene and AR protein expression, will be required to firmly establish this. Although further analysis of our tumours was constrained by the availability and nature of the FFPE tumour material, it is plausible that mutations that effect AR expression exist outside the promoter region examined, particularly in regions upstream of the AR 5'UTR and in the AR 3'UTR, which are reported to contain putative regulatory elements involved in controlling mRNA stability (Yeap et al., 2004). A recent epidemiological meta-analysis of the AR gene concluded that common polymorphisms in the AR gene are not associated with breast cancer risk among Caucasian women (Cox et al., 2006). However, the functional significance of these variants, the AR expression status and survival outcome of these patients was not considered in this study.

miR-124 does not regulate the AR 3'UTR in breast cancer cells.

MiRNAs are small non-coding RNAs of ~20nt in length that are capable of modulating gene expression post-transcriptionally. Many have been shown to act as either oncogenes or tumour suppressor genes that are crucial to the development of breast cancer metastasis and survival outcome (Hurst et al., 2009a). To address the possibility that altered expression of AR is mediated by differential expression of miRNAs, we screened the AR 3'UTR for potential miRNA target sites. Bioinformatic analysis revealed that miR-124 was the only miRNA predicted to target the human AR 3'UTR using miRanda and TargetScan. To examine whether miR-124 regulates the expression of the AR transcript, we used a reporter gene assay. MCF7 and T47D cells were transfected with pcDNA 3.1(+)-mir-124 vector and expression of miR-124 verified. Cells transfected with pcDNA 3.1(+)-mir-124 expressed high levels of mature miR-124 at 12, 24, 39 and 48 hr time intervals post-transfection, whereas no endogenous expression was detected in control-transfected cells (Fig. 6a). For luciferase assays, we co-transfected MCF7 and T47D cells with pcDNA 3.1(+)-mir-124 vector with the pSG5-AR 3'UTR vector. The introduction of the AR 3'UTR into pSG5 luc significantly increased reporter activity in MCF7 ($P=0.007$) but not in T47D cells (Fig. 6b). miR-124 overexpression did not alter luciferase activity of the AR 3'UTR construct in any of the cell lines examined (Fig. 6b). These results suggest that miR-124 is unlikely to regulate AR expression in these cells.

MiR-124 was the only miRNA predicted by two commonly used algorithms to target the AR 3'UTR. However, over-expression of miR-124 did not regulate AR expression in MCF7 or T47D cells, which otherwise display no endogenous expression of this miRNA. This suggests that either miR-124 is unable to regulate AR in this particular experimental system, or that the prediction not correct. There is certainly evidence suggesting that although a plethora of targets are predicted for miRNAs, that many of these are false positives (Valastyan et al., 2009a). There have been reports of other miRNAs regulating the expression of AR, including miR-488 (Sikand et al., 2010). It would be of interest to determine whether there is an association between the levels of these miRNAs and AR in Stage III breast cancer, and whether these miRNAs have the potential to be informative biomarkers or therapeutic targets for this disease. Interestingly, the inclusion of the AR 3'UTR significantly increased reporter activity in MCF7 but not in T47D cells suggesting the potential cell specific importance of this region in the post-transcriptional regulation in breast cancer cells.

Conclusions

In this paper we show that AR expression is significantly associated with 10-year survival outcome in patients with Stage III breast cancer. To predict and potentially address the poor survival outcome of patients with AR negative breast tumours it is important to understand the mechanism underlying the reduced AR expression. Here we demonstrate for the first time that hypermethylation of sections of the AR 5' regulatory region is associated with loss of AR expression in breast cancer cell lines, although not in a small set primary tumours. We describe a new somatic mutation in the AR 5'UTR which is found in two independent tumours and is not a normal polymorphism. This study highlights the potential for AR expression to be an informative prognostic biomarker for breast cancer survival and sets the scene for a more comprehensive investigation of the molecular basis of this phenomenon.

Methods

Breast cancer cell lines

Breast cancer cell lines MDAMB157 (ER-PR-), MDAMB231 (ER-PR-HER2-), MDAMB436 (ER-PR-), MDAMB468 (ER-PR-), MCF7 (ER+PR+HER2-), T47D (ER+PR+HER2-), ZR75-1 (ER+PR-), Hs578T (ER-PR-HER2-), and BT549 (ER-PR-) were obtained from American Type Culture Collection (ATCC) and cultured according to the manufacturer's recommendations. Hormone receptor status sourced from Neve et al, 2006.

Clinical samples

Primary breast tumours were sourced from the Princess Alexandra Hospital (Brisbane, Australia) following procedures endorsed by both local and national Human Ethics committees. Tumour tissues were formalin fixed and paraffin embedded (FFPE), sectioned and stained using hematoxylin and eosin (H&E), and confirmed by a qualified pathologist (G.D.F.) as invasive Grade III ductal adenocarcinomas. After the exclusion of patients for whom there was no follow-up data, non-breast cancer associated deaths and unreadable IHC expression, 73 patients were available for further analysis. None of the patients had received preoperative radiochemotherapy. The characteristics of the patients and tumours is shown in Table 1, with the following definitions of breast cancer stage III:

Stage IIIA: No tumor is found in the breast. Cancer is found in axillary lymph nodes that are attached to each other or to other structures, or cancer may be found in lymph nodes near the breastbone; or the tumor is 2 centimeters or smaller. Cancer has spread to axillary lymph nodes that are attached to each other or to other structures, or cancer may have spread to lymph nodes near the breastbone; or the tumor is larger than 2 centimeters but not larger than 5 centimeters. Cancer has spread to axillary lymph nodes that are attached to each other or to other structures, or cancer may have spread to lymph nodes near the breastbone; or the tumor is larger than 5 centimeters. Cancer has spread to axillary lymph nodes that may be attached to each other or to other structures, or cancer may have spread to lymph nodes near the breastbone.

Stage IIIB: In stage IIIB, the tumor may be any size and cancer: has spread to the chest wall and/or the skin of the breast; and may have spread to axillary lymph nodes that may be attached to each other or to other structures, or cancer may have spread to lymph nodes near the breastbone.

Cancer that has spread to the skin of the breast is inflammatory breast cancer. See the section on Inflammatory Breast Cancer for more information.

Stage IIIC: In stage IIIC, there may be no sign of cancer in the breast or the tumor may be any size and may have spread to the chest wall and/or the skin of the breast. Also, cancer: has spread to lymph nodes above or below the collarbone; and may have spread to axillary lymph nodes or to lymph nodes near the breastbone.
Cancer that has spread to the skin of the breast is inflammatory breast cancer. See the section on Inflammatory Breast Cancer for more information.
Stage IIIC breast cancer is divided into operable and inoperable stage IIIC.
In operable stage IIIC, the cancer: is found in ten or more axillary lymph nodes; or is found in lymph nodes below the collarbone; or is found in axillary lymph nodes and in lymph nodes near the breastbone.
In inoperable stage IIIC breast cancer, the cancer has spread to the lymph nodes above the collarbone.

Tissue microarray blocks and immunohistochemical staining
Tumour-rich tissue from each biopsy was distinguished from surrounding normal tissue in H&E-stained sections by a qualified pathologist (G.D.F). Tissue microarrays were constructed in duplicate from tumour-rich tissue cores (1mm x 0.6mm) using an automated tissue microarray (TMA) instrument (ATA-27; Beecher Instruments). 4µM sections of the TMA blocks were used for immunohistochemical (IHC) analysis. Sections were transferred on to glass slides, deparaffinised and immunostained using anti-ER (SP1), anti-PR, (SP2), anti-HER2/neu (4B5) (Ventana Medical Systems, pre-diluted) or anti-AR (Biocare Medical, 1:50 dilution) antibodies, and counterstained with 3,3'-Diaminobenzidine (DAB) and hematoxylin. Staining was performed with the BenchMark automated slide stainer (Ventana) using the iVIEW DAB detection kit with additional Avidin and Biotin Blockers according to the manufacturer's instructions. Analysis of stained sections was performed by a qualified pathologist (G.D.F) and the presence of tumour tissue confirmed by examining the counterstain. Expression was scored as positive when visible staining $\geq$ 1% was observed in the nucleus for ER, PR and AR. For HER2 IHC was semiquantitatively evaluated with a score of 3+ regarded as positive, 2+ as equivocal, and 1+ or 0 as negative. In instances of an equivocal evaluation, silver-enhanced in situ hybridisation was performed as previously described (Francis et al., 2009b).

DNA methylation analysis

Genomic DNA was isolated from cell lines using the NucleoSpin Tissue kit (MachereyNagel) according to the manufacturer's instructions. For each human tumour sample, four FFPE tumour-rich tissue cores (1mm x 0.6mm) were crushed and digested with proteinase K at 55°C for 2 days and treated with 20mg RNase A for 1 hr at 37°C. DNA was isolated using the PureGene kit (Qiagen) and subjected to bisulfite modification using the MethylEasy Xceed kit (Human Genetic Signatures) according to the manufacturer's instructions.

PCR amplification and methylation sensitive high resolution melt analysis (MS-HRM) was performed in duplicate on the RotorGeneTM Q (Qiagen). Primers were designed according to the principles outlined in (Wojdacz et al., 2008) to control for PCR bias and are shown in Supplementary Table 1. PCR was performed using 2ng of bisulphite modified template and standard PCR conditions, followed by one cycle of 1min 30sec at 72°C and an MS-HRM step from 70°C to 90°C rising by 0.1°C/sec. Bisulfite treated CpGenomeTM Universal Methylated DNA (Chemicon, Millipore) and DNA from T47D were used as positive/methylated and negative/unmethylated controls, respectively. The methylation status of these controls was confirmed by direct sequencing of MS-HRM products, purified using the QiaQuick Gel Extraction Kit (Qiagen), performed by the Australian Genome Research Facility (AGRF, Brisbane, Australia). To create a range of methylated standards, these controls were mixed in 25, 50 and 75% methylated to unmethylated template ratios and were included in the analysis of each region.

Sequenom MassARRAY DNA Methylation Analysis

Sequenom MassARRAY methylation analysis was performed as described previously (Coolen et al., 2007). The forward primer has a 10-mer tag (5-aggaagagag-3) and the reverse primer has a T7-promoter tag (5-cagtaatacgactcactatagggagaaggct-3) (Supplementary Table 1). Bisulfite treated CpGenomeTM Universal Methylated DNA (Chemicon, Millipore) and Whole genome amplified DNA prepared as per instructions with the GenomePlex® Complete Whole Genome Amplification kit (Sigma) were used as positive/methylated and negative/unmethylated controls, respectively. Triplicate PCR reactions were pooled and Shrimp Alkaline Phosphatase (Sequenom, San Diego) treatment performed followed by transcription and RNaseA Cleavage for the T-cleavage reaction. Purified samples were nanodispensed onto silicon chips preloaded with matrix

(SpectroCHIPs; Sequenom, San Diego). Mass spectra were collected using a MassARRAY mass spectrometer (Bruker-Sequenom) and results analysed by the EpiTYPER software V 1.0. Methylation readings with overlapping signals and silent peaks were eliminated from the calculation.

Quantitation of AR mRNA

To quantitate AR mRNA from cell lines, total RNA was extracted using Trizol (Invitrogen). cDNA was synthesised using 500ng of RNA and Superscript First Strand Synthesis System III (Invitrogen), according to the manufacturer's instructions. ß-actin was used to normalise mRNA concentration and primers are shown in Supplementary Table 1. Real-time PCR was performed in duplicate using the RotorGeneTM Q (Qiagen) using 50 cycles of standard PCR conditions.

Somatic mutation analysis

In breast cancer cell lines, a region spanning -659 to +280 with respect to the start of transcription (+1) were examined for somatic mutations. In FFPE tumours, the fragmented nature and limited availability of DNA meant that analysis was constrained to a smaller region (-165 to +133) and was examined in the 32 tumours for which IHC indicated negative AR expression. Primer sequences are shown in Supplementary Table 1. PCR was performed using KAPAHiFi DNA polymerase (KAPA Biosystems, Geneworks, Australia) and 50ng of template using 30 amplification cycles as per the manufacturer's instructions. PCR reactions were purified using the QiaQuick Gel Extraction Kit (Qiagen) and sequencing performed by AGRF (Brisbane, Australia).

Transcription factor binding site analysis

Bioinformatic analysis initially involved an analysis of UCSC Genome Browser (http://genome.ucsc.edu). Transcription factor binding sites were predicted by MOODS (MOtif Occurrence Detection Suite) (Korhonen et al., 2009). MOODS uses the standard scoring model (log-odds against the background distribution) of PWMs. Scoring thresholds were specified by *P*-value less than or equal to 0.01. We tested T105A 5'UTR variant and WT sequences for overlap with the *TFBS* models in the *JASPAR* database (Bryne et al., 2008). TFBS logos were downloaded from the JASPAR database web server: http://jaspar.cgb.ki.se/.

AR 5'UTR reporter assays

The AR 5'UTR T105A mutation was introduced into the wild-type AR 5'UTR sequence (1116 bp, GenBank Accession No. NG009014.1) by site-directed mutagenesis using a two-step PCR procedure using the primers listed in Supplementary Table 1 and standard PCR conditions. The AR 5'UTR wild-type and T105A mutant alone and together with the AR promoter were cloned into the HindIII/NcoI sites of pGL3-promoter and the KpnI/NcoI sites of pGL3-basic (Promega), respectively. Constructs were confirmed by sequencing performed by AGRF (Brisbane, Australia).

MCF7 and T47D cells were transiently transfected with 800ng pGL3+/- either the SV40 or the AR promoter, together with either the wild-type or mutant AR 5'UTR sequence and 100ng Renilla reporter in a 24-well plate, using Lipofectamine 2000 (Invitrogen). Forty-eight hours after the initial transfections, relative luciferase activities were determined using the Dual-Glo luciferase assay kit (Promega) and a DTX880 Multimode Detector (Beckman Coulter) according to the manufacturer's instructions. Statistical analysis was performed using unpaired, two-tailed t tests, with p values <0.05 considered significant.

miRNA analysis

Two algorithms, miRanda and TargetScan 5.1, were used to predict target sites for miRNA in the AR 3'UTR (Miranda et al., 2006, Lewis et al., 2005). A total of 55 miRNA were predicted to target AR, 12 of which have conserved seed sequences. Only miR-124 was predicted by both algorithms. To quantitate the expression of miR-124, total RNA from transfected MCF7 and T47D cells was extracted using Trizol (Invitrogen) and the expression of miR-124 determined relative to RNU6B using the miScript PCR System according to the manufacturer's instructions (Qiagen).

AR 3'UTR reporter assays

The AR 3'UTR (436bp; GenBank Accession No. NG009014.1) was ligated downstream of the luciferase coding sequence in the pSG5 vector (Stratagene). The mir-124-1 stem-loop (84bp; miRBase Accession No. MI0000443) + 200bp of flanking sequences was cloned into the KpnI/XbaI sites of the pcDNA 3.1(+) expression vector (Invitrogen). MCF7 and T47D cells were transiently transfected with 400ng pcDNA 3.1(+) constructs, 200ng pSG5-luciferase constructs and 10ng of Renilla reporter in a 24-well plate, using FuGene 6 (Roche). Forty-eight hours after the initial transfections, relative luciferase activities were

determined using the Dual-Glo luciferase assay kit (Promega) and a DTX880 Multimode Detector (Beckman Coulter) according to the manufacturer's instructions. Statistical analysis was performed using unpaired, two-tailed t tests, with p values <0.05 considered significant.

Statistical analysis

Survival time was calculated from the date of tumour removal to the date of last follow-up or death attributable to breast cancer. Overall survival probabilities were estimated non-parametrically with the use of the Kaplan-Meier product limit method (GraphPad Prism version 5.0a) and statistical significance accepted as $P \leq 0.05$.

Competing interests

The authors declare no competing interests.

Authors' contributions

Kate M Peters[1]: Contributed to AR IHC analysis, developed AR MS-HRM assay, determined expression levels in Breast cancer cell-lines. Drafted manuscript.

Stacey L Edwards[1]: Performed AR 5'UTR reporter assays

Shalima S Nair[2]: Performed AR Sequenom methylation assays on clinical samples and cell-lines

Juliet D French[1]:

Peter J Bailey[1]: performed bioinformatics analysis of AR 5'UTR and promoter region

Kathryn Salkield[1]: Contributed to AR IHC analysis

Sandra Stein[3]: Managed breast tumour biobank, created tumour tissue array, optimized AR antibody IHC, performed AR, ER, PR and HER2 IHC

Sarah Wagner[3]: Contributed to creation of tumour tissue array and optimization of AR antibody IHC

Glenn D Francis[3]: Pathologist responsible to tumour bank, pathological review of tumour samples and analysis of ER, PR, HER2 and AR IHC results

Susan J Clark[2]: Contributed to project design, attracting research funding, interpreting results and review of manuscript

Melissa A Brown[1] Project leader. Contributed to project design, attracting research funding, interpreting results and reviewing, editing, submission and post-review revision of manuscript.

Acknowledgements

This project was funded by grants from the National Breast Cancer Foundation (NBCF), Cancer Australia and National Health and Medical Research Council (NHMRC). SE was supported by an NHMRC CJ Martin Fellowship.

Figure legends

Fig. 1 Schematic diagram of the human AR gene. The relative positions of the two transcription initiation sites (TIS I and II) and functionally known motifs; CpG islands, cAMP responsive element (CRE), helix-loop-helix-like (HL) motifs, a palindromic homopurine repeat (grey box) and GC-box (black box) which contains an SP1 binding site, are indicated (GenBank Accession No. NG009014.1). A grey circle denotes the T105A alteration identified in two primary breast tumours in the present study. DNA methylation was assessed in Regions 1-4, as denoted by black lines. UTR, untranslated region. ATG, translation start site.

Fig. 2 Impact of Androgen Receptor expression on breast cancer survival . Kaplan-Meier estimates of 10 year survival of patients by AR expression for primary tumours. n denotes the number of patient samples.

Fig. 3 DNA methylation of the AR gene is associated with loss of AR protein expression in breast cancer cells. (a) DNA methylation status of breast cancer cell lines. Three regions of AR were assessed by MS-HRM, overlapping Regions 1 and 2 in the AR minimal promoter and Region 3 at the translational start site (refer to Fig. 1). Region 4 was assessed by Sequenom MassARRAY. Data represent the average of two independent experiments. (b) AR mRNA expression in breast cancer cell lines was assessed by qPCR. Expression is shown relative to ß-actin and bars represent the mean + standard deviation of two independent experiments.

Fig. 4 DNA methylation of an 115bp region in the AR promoter is not associated with loss of AR protein expression in a selective group of primary breast tumours. Methylation of six CpG dinucleotides in the AR promoter of AR positive (Pos) and negative (Neg) Grade III breast tumours as determined by Sequenom MassARRAY (Region 4, Fig. 1). Due to cleavage patterns, the average methylation was determined for CpG dinucleotides 2 and 3. Each dot represents an individual patient. A methylation value of 1.0 indicates a fully methylated amplicon, while a value of 0.0 indicates a fully unmethylated amplicon. Horizontal lines represent average cohort methylation.

Fig. 5 Impact of the AR T105A 5'UTR variant. Luciferase reporter activities of pGL3 basic with the AR promoter (AR prom) and pGL3 promoter vector which contains the SV40 promoter (SV40 prom) together with either the wild-type (WT) or T105A mutant (MT) AR 5'UTR sequence in MCF7 and T47D cells. Data is shown relative to the respective empty vector + the standard error of the mean (SEM) and was generated from three independent experiments. ***P=0.0001.

Fig. 6 miR-124 does not target the AR 3'UTR in breast cancer cells. (a) miR-124 expression in MCF7 and T47D cells was assessed by qPCR following transfection with pcDNA 3.1(+)-mir-124. Expression is shown relative to RNU6B and dots represent the mean + standard deviation of two independent experiments. (b) Luciferase reporter activities relative to the Renilla internal control + the standard error of the mean (SEM) are shown. Data were generated from three independent experiments. **P=0.007.

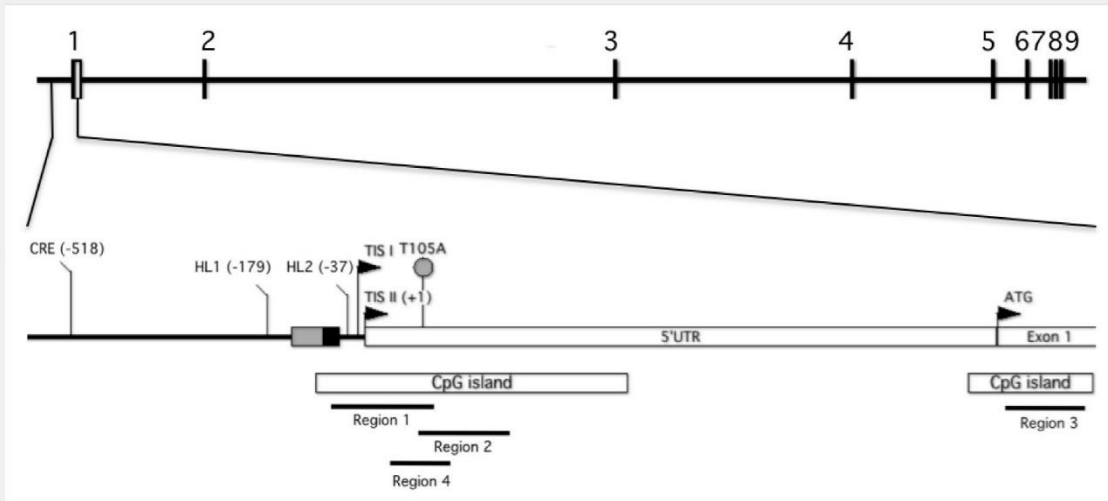


Figure 1.

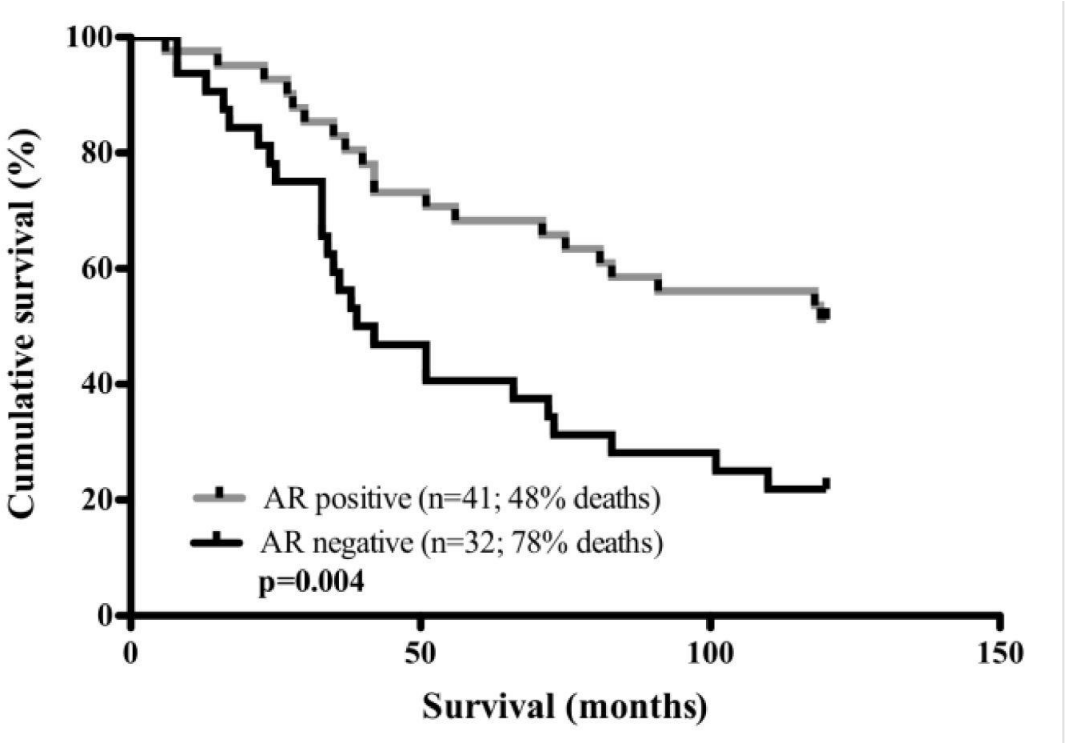

100
80
60
40
20
0
Cumulative survival (%)
AR positive (n=41; 48% deaths)
AR negative (n=32; 78% deaths)
p=0.004
0
50
100
150
Survival (months)

Figure 2.

a

	Region 1	Region 2	Region 3	Region 4
# CpG's assessed	9	14	8	6
Cell line				
Universally Methylated				
MDAMB231				
MCF7				
MDAMB157				
MDAMB468				
MDAMB436				
Hs578T				
BT549				
ZR751				
T47D				

DNA methylation
0-25%
26-49%
50-74%
75-100%

b

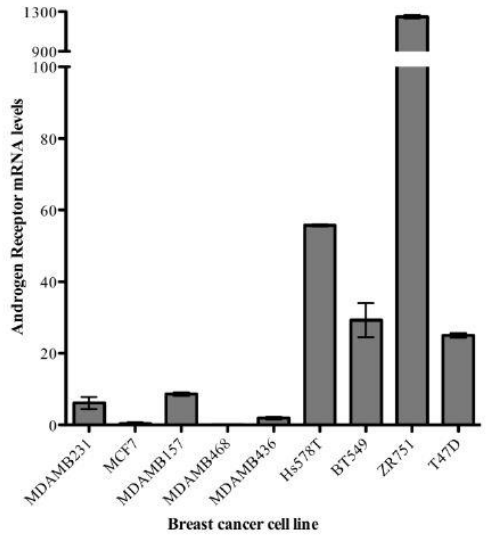


Figure 3.

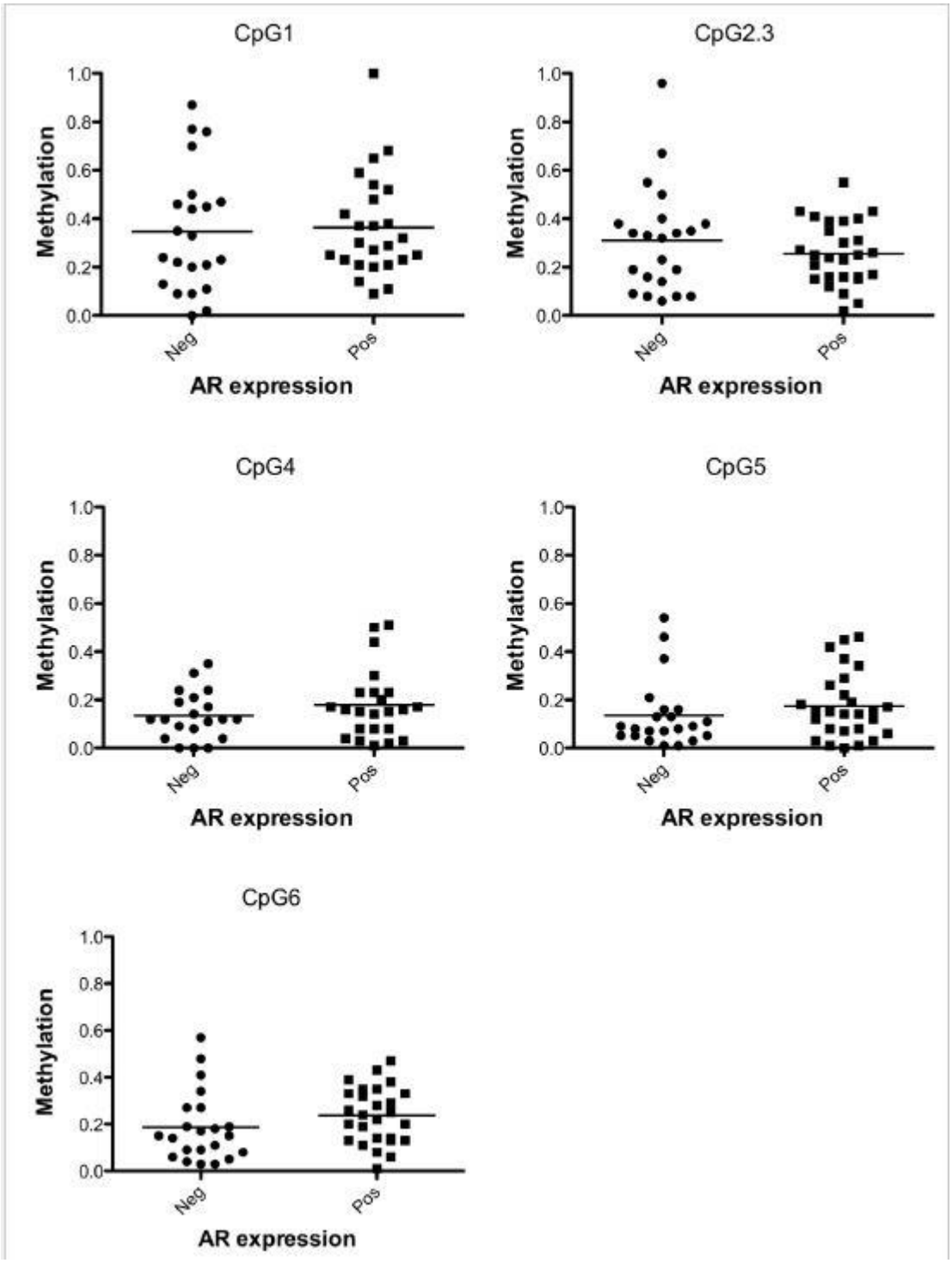

CpG1
CpG2.3
CpG4
CpG5
CpG6
Methylation
1.0
0.8
0.6
0.4
0.2
0.0
Neg
Pos
AR expression

Figure 4.

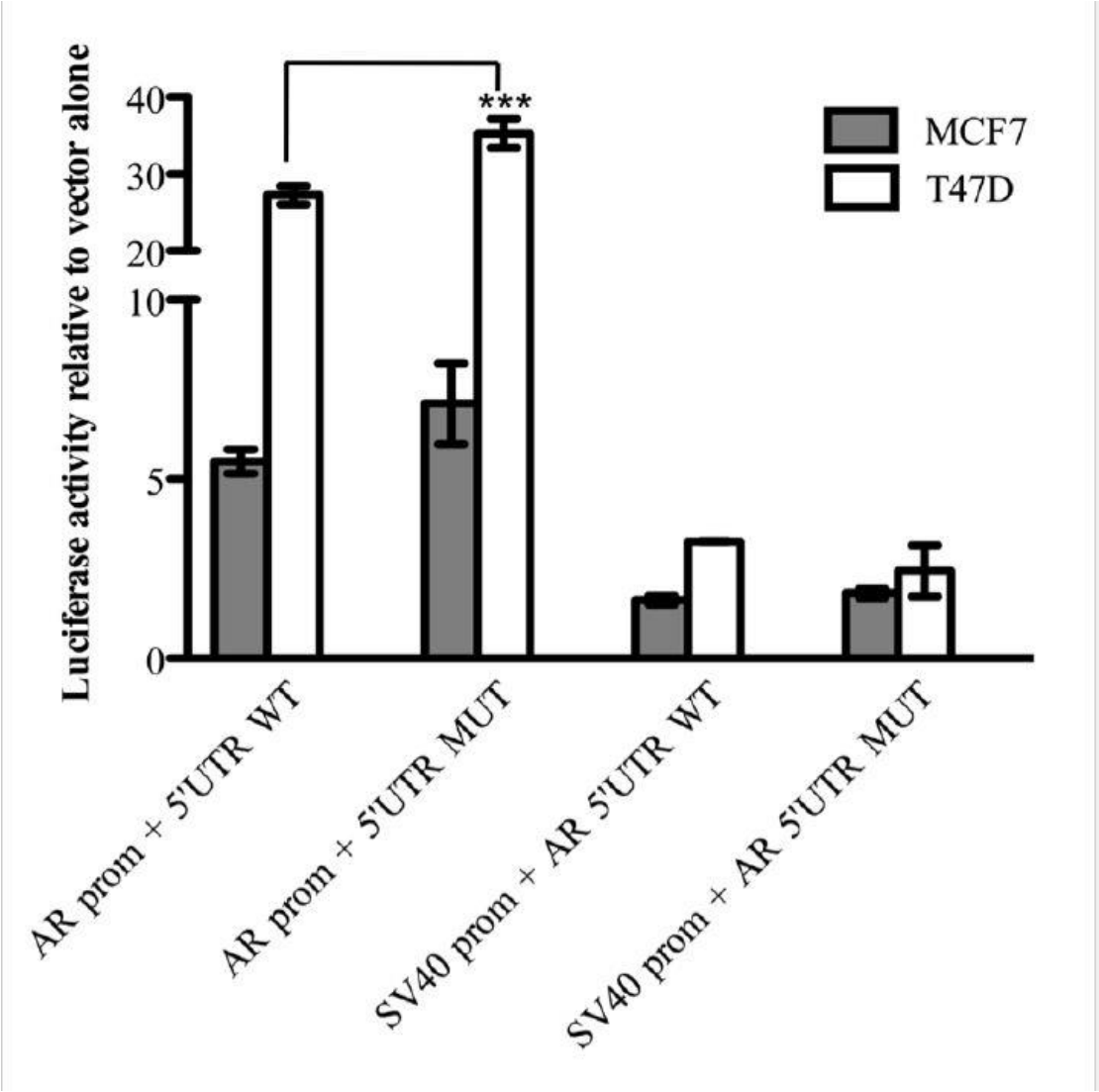

Luciferase activity relative to vector alone
40
30
20
10
5
0
***
MCF7
T47D
AR prom + 5'UTR WT
AR prom + 5'UTR MUT
SV40 prom + AR 5'UTR WT
SV40 prom + AR 5'UTR MUT

Figure 5.

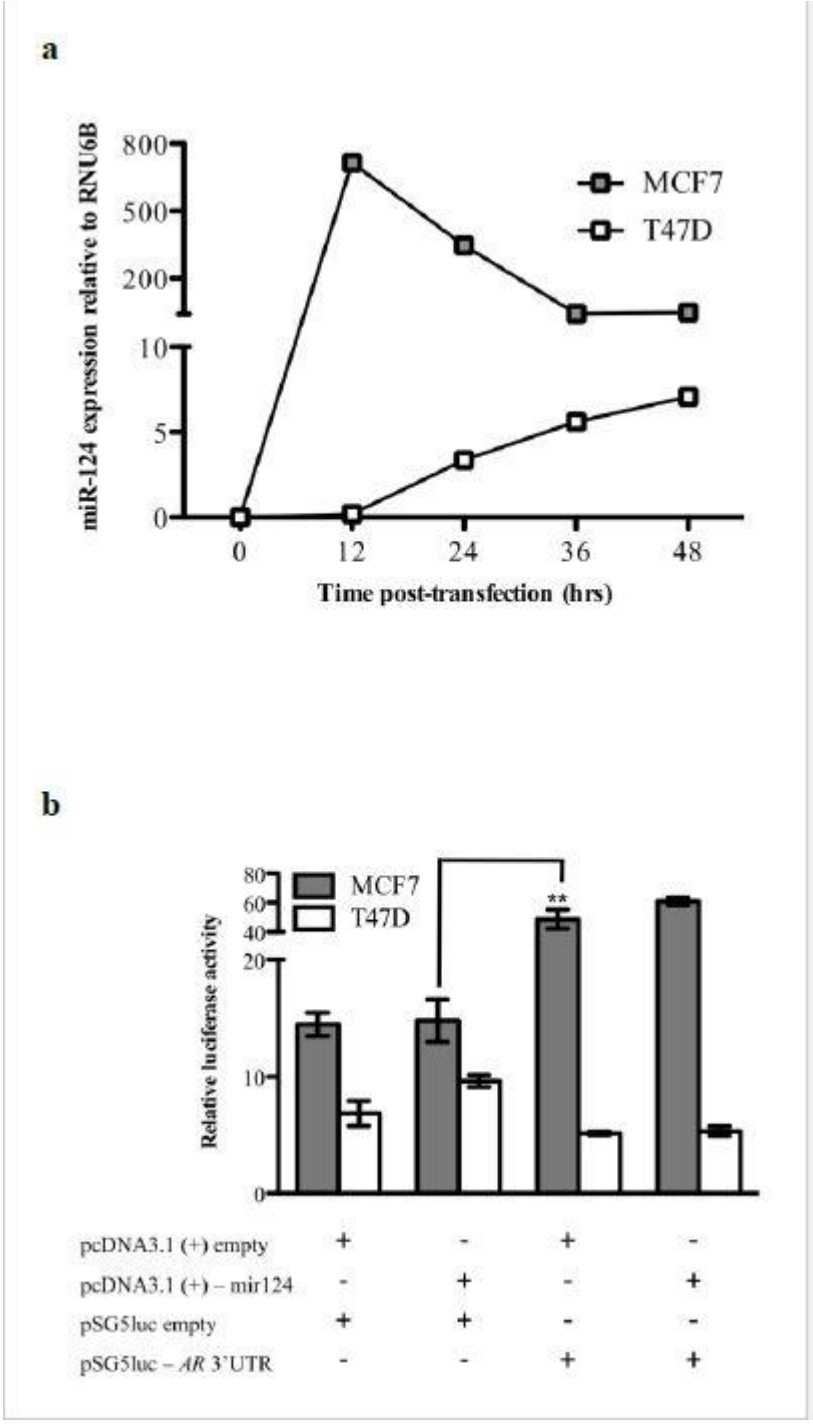


Figure 6.

Table 1 Patient and tumour characteristics*

Factor	AR negative n=32 (%)	AR positive n=41 (%)
Age (years)		
≤35	1 (3.0)	1 (2.4)
≥35	31 (97.0)	40 (97.6)
Estrogen receptor		
Negative	23 (72.0)	9 (22.5)
Positive	9 (28.0)	31 (77.5)

Progesterone receptor		
Negative	27 (87.0)	16 (39.0)
Positive	4 (13.0)	25 (61.0)
HER2		
Negative	18 (85.7)	25 (75.8)
Positive	3 (14.3)	8 (24.2)
Triple negative[a]		
No	6 (28.6)	30 (90.9)
Yes	15 (71.4)	3 (9.1)

*All patients were diagnosed with Stage III disease, as defined in the Materials and Methods section of the manuscript

[a]Triple negative breast cancer represents tumours displaying negative expression for estrogen receptor, progesterone receptor and HER2 by IHC. AR, androgen receptor; HER2, human epidermal growth factor receptor type 2.

References

1. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS et al: Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A 2001, 98(19):10869-10874.

2. Rakha EA, Reis-Filho JS, Ellis IO: Combinatorial biomarker expression in breast cancer. Breast Cancer Res Treat 2010, 120(2):293-308.

3. Yeh S, Hu YC, Wang PH, Xie C, Xu Q, Tsai MY, Dong Z, Wang RS, Lee TH, Chang C: Abnormal mammary gland development and growth retardation in female mice and MCF7 breast cancer cells lacking androgen receptor. J Exp Med 2003, 198(12):1899-1908.

4. Liao DJ, Dickson RB: Roles of androgens in the development, growth, and carcinogenesis of the mammary gland. J Steroid Biochem Mol Biol 2002, 80(2):175-189.

5. Li S, Han B, Liu G, Ouellet J, Labrie F, Pelletier G: Immunocytochemical localization of sex steroid hormone receptors in normal human mammary gland. J Histochem Cytochem 2010, 58(6):509-515.

6. Doane AS, Danso M, Lal P, Donaton M, Zhang L, Hudis C, Gerald WL: An estrogen receptor-negative breast cancer subset characterized by a hormonally regulated transcriptional program and response to androgen. Oncogene 2006, 25(28):3994-4008.

7. Birrell SN, Bentel JM, Hickey TE, Ricciardelli C, Weger MA, Horsfall DJ, Tilley WD: Androgens induce divergent proliferative responses in human breast cancer cell lines. J Steroid Biochem Mol Biol 1995, 52(5):459-467.

8. Ni M, ., Chen Y, Lim E, Wimberly H, Bailey ST, Imai Y, Rimm DL, Liu XS, Brown M: Targeting androgen receptor in estrogen receptor-negative breast cancer. Cancer Cell 2011, 20:119-131.
9. Agoff SN, Swanson PE, Linden H, Hawes SE, Lawton TJ: Androgen receptor expression in estrogen receptor-negative breast cancer. Immunohistochemical, clinical, and prognostic associations. Am J Clin Pathol 2003, 120(5):725-731.
10. Gonzalez-Angulo AM, Stemke-Hale K, Palla SL, Carey M, Agarwal R, Meric-Berstam F, Traina TA, Hudis C, Hortobagyi GN, Gerald WL et al: Androgen receptor levels and association with PIK3CA mutations and prognosis in breast cancer. Clin Cancer Res 2009, 15(7):2472-2478.
11. Faber PW, van Rooij HC, van der Korput HA, Baarends WM, Brinkmann AO, Grootegoed JA, Trapman J: Characterization of the human androgen receptor transcription unit. J Biol Chem 1991, 266(17):10743-10749.
12. Faber PW, van Rooij HC, Schipper HJ, Brinkmann AO, Trapman J: Two different, overlapping pathways of transcription initiation are active on the TATA-less human androgen receptor promoter. The role of Sp1. J Biol Chem 1993, 268(13):9296-9301.
13. Chen S, Supakar PC, Vellanoweth RL, Song CS, Chatterjee B, Roy AK: Functional role of a conformationally flexible homopurine/homopyrimidine domain of the androgen receptor gene promoter interacting with Sp1 and a pyrimidine single strand DNA-binding protein. Mol Endocrinol 1997, 11(1):3-15.
14. Takane KK, McPhaul MJ: Functional analysis of the human androgen receptor promoter. Mol Cell Endocrinol 1996, 119(1):83-93.
15. Mizokami A, Yeh SY, Chang C: Identification of 3',5'-cyclic adenosine monophosphate response element and other cis-acting elements in the human androgen receptor gene promoter. Mol Endocrinol 1994, 8(1):77-88.
16. Kinoshita H, Shi Y, Sandefur C, Meisner LF, Chang C, Choon A, Reznikoff CR, Bova GS, Friedl A, Jarrard DF: Methylation of the androgen receptor minimal promoter silences transcription in human prostate cancer. Cancer Res 2000, 60(13):3623-3630.
17. Waltering KK, Wallen MJ, Tammela TL, Vessella RL, Visakorpi T: Mutation screening of the androgen receptor promoter and untranslated regions in prostate cancer. Prostate 2006, 66(15):1585-1591.
18. Crocitto LE, Henderson BE, Coetzee GA: Identification of two germline point mutations in the 5'UTR of the androgen receptor gene in men with prostate cancer. J Urol 1997, 158(4):1599-1601.

19. Cox DG, Blanche H, Pearce CL, Calle EE, Colditz GA, Pike MC, Albanes D, Allen NE, Amiano P, Berglund G et al: A comprehensive analysis of the androgen receptor gene and risk of breast cancer: results from the National Cancer Institute Breast and Prostate Cancer Cohort Consortium (BPC3). Breast Cancer Res 2006, 8(5):R54.
20. Karlsson C, Bodin L, Piehl-Aulin K, Karlsson MG: Tissue Microarray Validation: A Methodologic Study with Special Reference to Lung Cancer. Cancer Epidemiology Biomarkers & Prevention 2009, 18:2014-2021.
21. Castellano I, Allia E, Accortanzo V, Vandone AM, Chiusa L, Arisio R, Durando A, Donadio M, Bussolati G, Coates AS et al: Androgen receptor expression is a significant prognostic factor in estrogen receptor positive breast cancers. Breast Cancer Res Treat 2010.
22. Bryan RM, Mercer RJ, Bennett RC, Rennie GC, Lie TH, Morgan FJ: Androgen receptors in breast cancer. Cancer 1984, 54(11):2436-2440.
23. Park S, Koo JS, Kim MS, Park HS, Lee JS, Kim SI, Park BW, Lee KS: Androgen receptor expression is significantly associated with better outcomes in estrogen receptor-positive breast cancers. Ann Oncol 2011.
24. Jarrard DF, Kinoshita H, Shi Y, Sandefur C, Hoff D, Meisner LF, Chang C, Herman JG, Isaacs WB, Nassif N: Methylation of the androgen receptor promoter CpG island is associated with loss of androgen receptor expression in prostate cancer cells. Cancer Res 1998, 58(23):5310-5314.
25. Coolen MW, Statham AL, Gardiner-Garden M, Clark SJ: Genomic profiling of CpG methylation and allelic specificity using quantitative high-throughput mass spectrometry: critical evaluation and improvements. Nucleic Acids Res 2007, 35(18):e119.
26. Yeap BB, Wilce JA, Leedman PJ: The androgen receptor mRNA. Bioessays 2004, 26(6):672-682.
27. Hurst DR, Edmonds MD, Welch DR: Metastamir: the field of metastasis-regulatory microRNA is spreading. Cancer Research 2009, 69(19):7495-7498.
28. Valastyan S, Benaich N, Chang A, Reinhardt F, Weinberg RA: Concomitant suppression of three target genes can explain the impact of a microRNA on metastasis. Genes Dev 2009, 23(22):2592-2597.
29. Sikand K, Slaibi JE, Singh R, Slane SD, Shukla GC: miR 488* inhibits androgen receptor expression in prostate carcinoma cells. Int J Cancer 2010.
30. Francis GD, Jones MA, Beadle GF, Stein SR: Bright-field in situ hybridization for HER2 gene amplification in breast cancer using tissue microarrays: correlation between

chromogenic (CISH) and automated silver-enhanced (SISH) methods with patient outcome. Diagn Mol Pathol 2009, 18:88-95.
31. Wojdacz TK, Hansen LL, Dobrovic A: A new approach to primer design for the control of PCR bias in methylation studies. BMC Res Notes 2008, 1:54.
32. Korhonen J, Martinmäki P, Pizzi C, P. R, Ukkonen E: MOODS: fast search for position weight matrix matches in DNA sequences. Bioinformatics 2009, 25:3181-3182.
33. Bryne JC, Valen E, Tang MH, Marstrand T, Winther O, da Piedade I, Krogh A, Lenhard B, Sandelin A: JASPAR, the open access database of transcription factor-binding profiles: new content and tools in the 2008 update. . Nucleic Acids Res 2008(36):D102-106.
34. Miranda KC, Huynh T, Tay Y, Ang YS, Tam WL, Thomson AM, Lim B, Rigoutsos I: A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes. Cell 2006, 126(6):1203-1217.
35. Lewis BP, Burge CB, Bartel DP: Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell 2005, 120(1):15-20.

Chapter 7

FRANCIS, G. D., STEIN, S. R. & FRANCIS, G. D. Prediction of histologic grade in breast cancer using an artificial neural network. Neural Networks (IJCNN), The 2012 International Joint Conference on, 10-15 June 2012 2012. 1-5.

7.1.1. Introduction

NeuralWorks Predict® is an integrated tool for rapidly creating and deploying prediction and classification applications. It combines neural network technology with genetic algorithms, statistics, and fuzzy logic to automatically find optimal or near-optimal solutions for a wide range of problems. The program can be used for a variety of analysis and interpretation problems. One advantage of NeuralWorks Predict® is that it requires no prior knowledge of neural networks. The input data is analysed to identify appropriate transforms and this data is then partitioned into training and test sets. The relevant input variables are selected and the program then constructs, trains, and optimizes a neural network tailored to the problem. However, the parameters can be modified by the user with direct access to all key training and network parameters. NeuralWorks Predict® runs as an add-in for Microsoft Excel to take advantage of Excel's rich data handling and graphing capabilities.

NeuralSight® extends the power and flexibility of the NeuralWorks Predict® neural network engine by incorporating sophisticated yet easy-to-use facilities for building and evaluating hundreds of neural network models with minimal effort and intervention. Features in NeuralSight® include Self-Organizing Map (SOM) clustering and visualization capabilities.

These programs were used to develop a model to predict grade in breast cancer using five IHC results (BUB1B, RaCGAP1, RRM2, NEK2,CENPA (Ma et al., 2008)).

Histologic grade in breast cancer is based on three morphologic features:

- Tubule formation
- Nuclear pleomorphism
- Mitotic count

A value of 1-3 is assigned to each parameter and the values are combined to produce a score:

- 3-5 = Grade 1
- 6 & 7 = Grade 2
- 8 & 9 =Grade 3

Molecular techniques have been used to stratify Grade 2 breast cancers into Grade 1 and grade 3 tumours as this appears to more accurately reflect patient outcome (Ignatiadis and Sotiriou, 2008a, Ivshina et al., 2006).

7.1.2. Results

Abstract — Histological grade is a historically used and well-documented prognostic indicator in breast cancer. There are three categories of grade (G1, G2 and G3) based on the degree of tubule formation, nuclear pleomorphism and mitotic count. A number of studies have reported that histological assessment is not uniformly reported. As a result of low inter-pathologist correlation associated with pathological diagnosis and non-standardised grading systems, patients are not always allocated into the correct grouping, and G2 has often been considered a “safe” group if one is unsure. A previously published study used real-time polymerase chain reaction (RT-PCR) for 5 genes to molecularly classify the G2 tumours into either G1 or G3. Due to the workflow constraints within pathology laboratories it was not considered feasible to molecularly profile every G2 tumour. In light of this we obtained the antibodies that corresponded to the 5 genes (BUB1B, CENPA, RACGAP1, RRM2 and NEK2) and performed immunohistochemistry (IHC) on formalin fixed paraffin embedded (FFPE) sections of 43 tumours (11 G1 and 32 G3). Results for all tumours were randomly divided into training and testing sets and an artificial neural network (NeuralSight and NeuralWare Predict) was used to classify the grade of tumours. Thirty-three additional G2 tumours were used for validation of the ANN. The ANN classified these tumours into 5 G1 and 28 G3 tumours. This predicted grade showed significant correlation with patient survival. Neural networks can be used to reclassify breast cancer G2 tumours into G1 and G3 using a panel of 5 IHC markers. This has the potential to impact on patient care, treatment decisions and outcome.
Keywords- Breast cancer, grade, artificial neural network.

I. INTRODUCTION

Breast cancer is the leading cause of cancer-related deaths in women world-wide [1] and is the second most common cancer in Australian women. More than 13,000 new cases are diagnosed each year in Australia [2], increasing from 11,000 in 2000 [3], of which 2,200 are in Queensland. Approximately 2,700 deaths from breast cancer occur annually in

Australia of which 465 were in Queensland in 2003. The number of hospital separations in Australia for women increased from 15,831 in 1995-96 to 23,598 in 2003-4 [2].

In current clinical practice for breast cancer, the standard prognostic factors that guide adjuvant systemic treatment decisions include tumour size, histologic subtype, histologic grade, oestrogen and progesterone receptor status, HER2 status and axillary lymph node status. A measure of the limitation of these factors in predicting outcome is exemplified by the observation that recurrence rates are approximately 25% in lymph node negative breast cancer patients who will often not receive adjuvant therapy [4].

Despite the published reports of new biomarkers to predict prognosis or response to therapy in breast cancer patients, only a limited number have entered clinical practice. A predictive factor is defined as a clinical or pathologic feature that determines the likelihood of a response to a particular treatment. A prognostic factor is defined as a clinical or pathologic biomarker that determines patient outcome or survival. A biomarker may be both predictive and prognostic. For a factor to be useful, it must be technically validated, clinically validated and influence clinical decision making. A large proportion of studies have used immunohistochemistry (IHC) to assess the expression of different antigens in breast cancer cells compared to normal tissue. Limitations of these studies include variation in patient selection, different primary antibody clones, different cut-off criteria, inter-laboratory variation and reproducibility of the methodology.

In addition, molecular techniques including DNA microarray analysis have indicated a large number of genes are involved in the progression of breast carcinoma. More than 3,000 genes have been implicated in distinguishing oestrogen receptor (ERp) positive tumours from ERp-negative tumours [5].

Breast cancer is graded according to the Elston-Ellis modified Scarff, Bloom and Richardson grading system also known as the Nottingham modified Bloom and Richardson system [6, 7]. There are three components assessed for assigning a grade to a tumour: tubule formation, nuclear pleomorphism and mitotic rate. Each component is assigned a number from 1 to 3 and the components are tallied to obtain a value out of 9. Tumours with a score of 5 or less are graded as Grade 1 (G1), scores of 6-7 are Grade 2 (G2) and scores of 8-9 as Grade 3 (G3) tumours. The grade of tumour shows significant correlation with breast cancer specific survival [8-11] for patients with lymph node metastases and those without nodal metastases at 5, 10 and 15 years (Fig. 1). The evaluation of tubule formation and mitotic rate also provide independent prognostic information [12].

Evaluation of large patient cohorts with multivariate analysis has shown that grade is a significant predictor of disease recurrence and patient death (Ellis et al., 1992, Schumacher et al., 1993, Roberti, 1997, Lundin et al., 2001). Untreated G1 patients have a 5 year survival rate of approximately 95%, whereas, those patients with G2 or G3 tumours have 5 year survival rates of approximately 75% and 50% respectively.

There is, however, significant inter-observer variability in the evaluation of histological grade by pathologists (Harvey et al., 1992, Robbins et al., 1995, Frierson et al., 1995, Dalton et al., 1994, Dalton et al., 2000) and the usefulness in patient prognosis has been questioned (Younes and Laucirica, 1997, Hayes et al., 2001). Grade 2 tumours comprise approximately 50% of all breast cancers and it has been proposed that refinement of this group may improve the prognostic value of tumour grading (Sotiriou et al., 2006, Ivshina et al., 2006).

Gene expression profiling has been used to more accurately grade tumours to overcome the limitations of the current system (Ivshina et al., 2006, Sotiriou et al., 2006). Sotiriou et al (Sotiriou et al., 2006) developed a 97-gene expression profile associated with histological grade and found that the gene expression profile correlated more strongly with relapse-free survival compared to histological grade (Sotiriou et al., 2006). Similarly, Ma et al, (Ma et al., 2003) and Ivshina et al, (Ivshina et al., 2006) have also developed histologic grade signatures and the latter enabled the refinement of the G2 tumours into different sub-grades with clinical implications (Ivshina et al., 2006). The Genomic Grade Index (Filho et al., 2010, Ignatiadis and Sotiriou, 2008b, Toussaint et al., 2009) has been further developed for both frozen and formalin fixed paraffin embedded tissue (FFPE). Ma et al, (Ma et al., 2008) developed a polymerase chain reaction (PCR) based five gene molecular grade index (MGI) for applications with FFPE material. The rationale for this approach was that it may be more easily implemented into a diagnostic laboratory. The MGI has been validated in a separate patient cohort (Jerevall et al., 2011).

In this study, the five gene panel of the MGI was used to develop IHC staining protocols for patient samples and an artificial neural network was used to develop an algorithm for determination of tumour grade.

Materials and Methods

Tissue microarrays (TMAs) were constructed using an automated Beecher ATA-27 and semi-automated Beecher Galileo (Beecher Instruments, Sun Prairie, USA). Areas of tumour were identified on haematoxylin and eosin (H & E) stained slides and tissue cores

were punched from these areas and inserted into recipient blocks. 4 µm sections were cut onto SuperFrost Plus slides and IHC was performed using primary antibodies against BUB1B, NEK2, CENPA, RacGAP1 and RRM2. Slides were digitally scanned on a Hamamatsu Nanozoomer (Hamamatsu City, Japan) at 40x magnification and attached to image maps in a database, SlidePath Distiller (Leica Microsystems, Wetzlar, Germany). The images were scored on-line using SlidePath Distiller. Each TMA core was scored for percentage of tumour staining (0-100%), intensity of tumour staining (0 – 3) and the localization in the cellular compartments; membrane, cytoplasmic or nuclear. Data, including morphologic grade, patient follow-up, survival and cause of death, was extracted from the database into Microsoft Excel (Microsoft Corporation, USA).

43 patients with complete data sets were randomly divided into training and teaching sets. Within this patient population there were 11 Grade 1 tumours and 32 Grade 3 tumours. NeuralSight (NeuralWare, Carnegie, PA) was used to evaluate 20 models through NeuralWare Predict (NeuralWare, Carnegie, PA). In NeuralSight, the Model Type selected in the program was Classification and build options included cascaded variable selection with noisy data and comprehensive and exhaustive variable selection options. In NeuralWare Predict, a multilayer perceptron neural network, the selected options included adaptive gradient learning rule, cascaded variable selection, an output layer function of SoftMax and an evaluation function of average classification rate. Cascaded variable selection was performed before the main variable step and was used to estimate the probability that a particular variable was present in a good solution, eliminating those variables that have a very low probability of inclusion in an optimum solution. Input variable selection used a genetic algorithm with multiple regression to select the model's input variables. The adaptive gradient learning rule used back-propagated gradient information to guide an iterative line search algorithm with weight decay used to avoid overfitting. The neural network architecture selected used 22 input nodes, 3 hidden nodes and 1 output node. NeuralWare Predict was used to run predicted grade and validate the algorithm on a separate data set of 33 Grade 2 tumours through a plugin in Microsoft Excel 2007. SPSS 19 (IBM Australia) was used for Kaplan-Meier survival curves for Breast Cancer Specific Survival (BCSS).

Results

Photomicrographs of morphologic G1 and G3 tumours are illustrated in Figure 2.

IHC staining showed variable intensity and percentage of cytoplasmic staining for BUB1B, RRM2 and NEK2 and variable intensity and percentage of nuclear staining for RacGAP1

and CENPA (Figure 3). Kaplan-Meier survival curves for BCSS for the tumours used in this study show differences in survival for the morphologic G1, G2 and G3 tumours with the poorest outcome for G3 tumours (Figure 4).

NeuralSight generated 20 models and these were evaluated using ROC modeling. Model 10 showed an ROC of 1 and this model was used to evaluate the G2 tumours. Of the 33 G2 tumours, 5 were predicted to be G1 and 28 were predicted to be G3 tumours. Kaplan-Meier survival curves showed that the predicted G1 tumours had a 100% long term BCSS, whereas, the predicted G3 tumours had a BCSS similar to morphologic G3 tumours (Figure 5).

Discussion

Breast cancer grade is a significant prognostic factor in breast cancer patients; however, morphologic assessment of grade is subjective and there is variability in interpretation between reporting pathologists. To overcome these limitations, molecular profiling has been used to develop algorithms to stratify G2 tumours into those tumours that are similar to morphologic G1 tumours and those similar to morphologic G3 tumours. The rational for this is that it will enable more accurate prognostic information to improve patient care and outcome.

Gene expression profiling and even real-time polymerase chain reaction (RT-PCR) are not easily integrated into a diagnostic anatomical pathology laboratory and for this reason, based on a small gene list, IHC was used to ascertain if it was feasible to use this technology to achieve the same goals.

This study shows that it is possible to use a panel of five IHC biomarkers, in conjunction with an artificial neural network, to stratify G2 breast cancers into two separate groups with differing survival outcomes. The neural network model in this application uses biomarkers that had previously been reported to be useful in predicting grade stratification. The parameters included a mixture of percentage of tumour cells staining and the intensity of staining which correlates with the protein expression of that particular biomarker. This format forms part of routine patient care in breast cancer for other biomarkers including oestrogen receptor, progesterone receptor and HER2. In addition, the design of model parameters enables variations in the interpretation of the IHC results to be accommodated by the network model. Further validation studies are in progress. If successful, this methodology has the potential to be readily integrated into the diagnostic laboratory to improve patient care.

Acknowledgment

The authors thank S. Leishman for formatting and proof-reading of the paper.

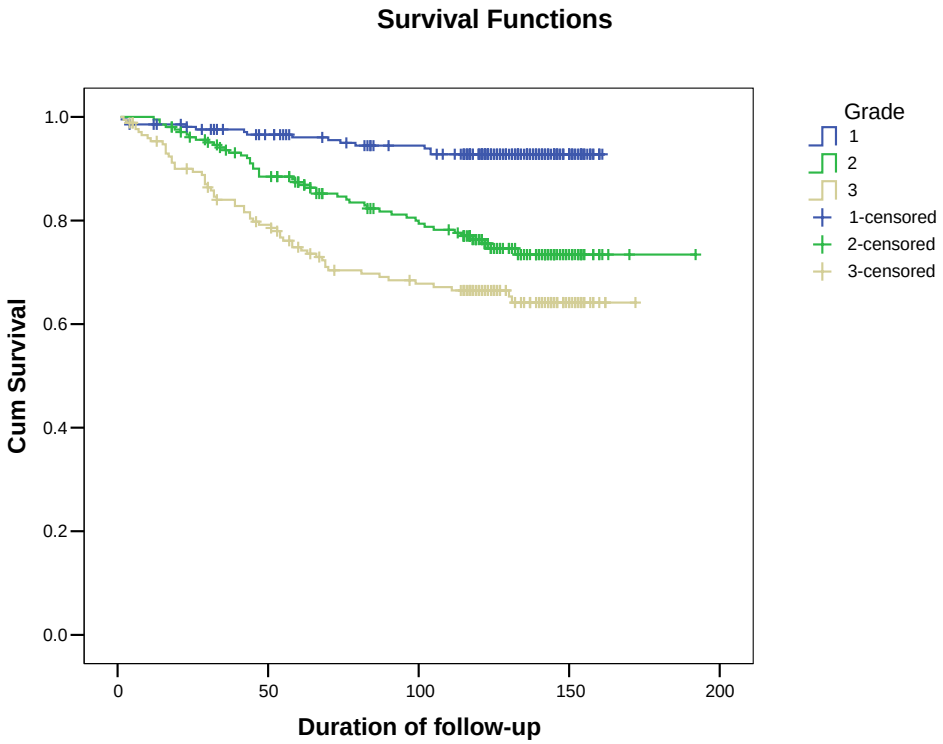


Figure 1: Kaplan-Meier survival curves for 591 breast cancer patients according to tumour grade over 200 months.

a) b)

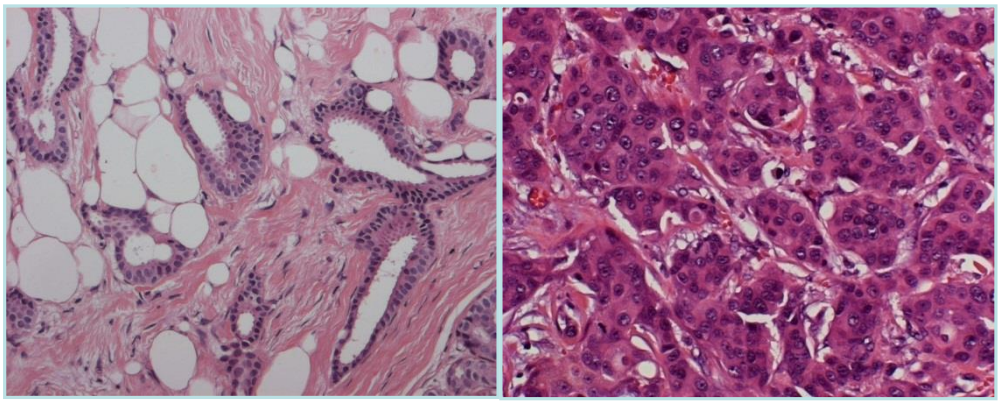


Figure 2: Haematoxylin and eosin stained sections illustrating (a) Grade 1 and (b) Grade 3 infiltrating ductal carcinoma (x20 magnification).

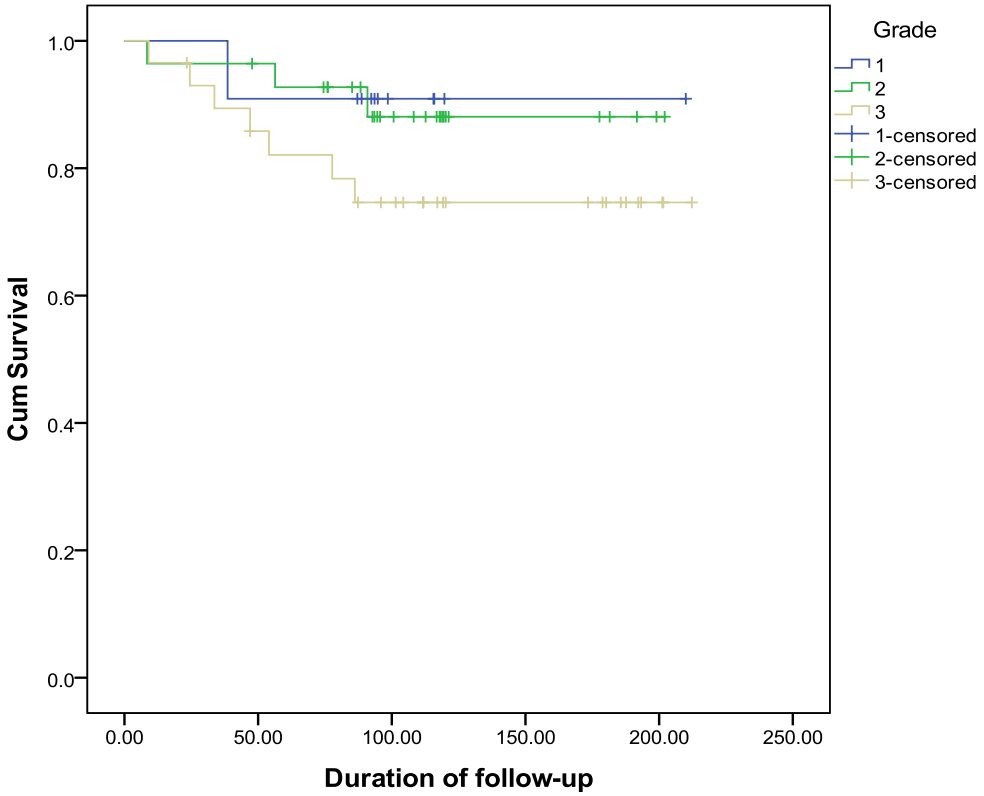


Figure 3: Kaplan-Meier BCSS survival curves for morphologic G1, G2 and G3 tumours over 250 months.

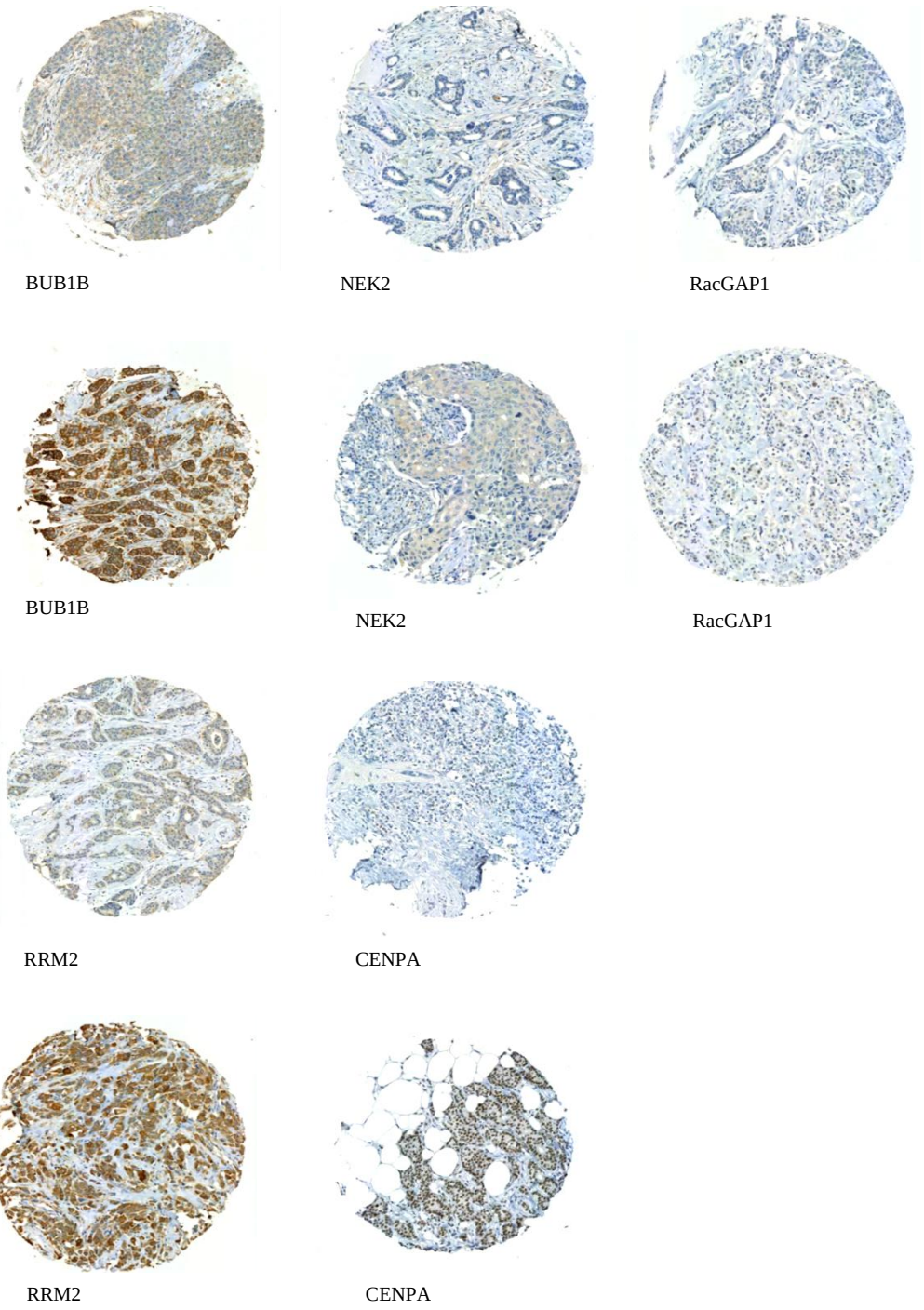


Figure 4: Immunohistochemical detection of BUB1B, NEK2, RacGAP1, RRM2 and CENPA in breast cancer TMAs (x20 magnification).

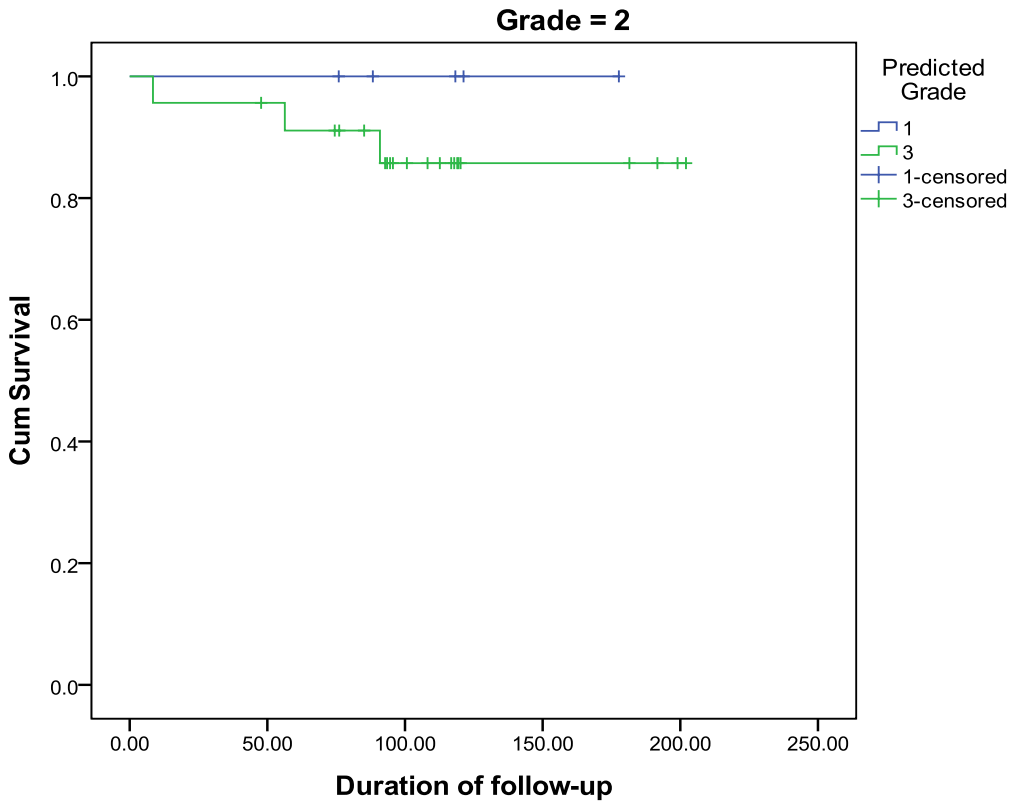


Figure 5: Kaplan-Meier BCSS survival curves for morphologic G2 tumours stratified into predicted G1 and G3 tumour over 250 months.

REFERENCES

Hinestrosa, M.C., et al., Shaping the future of biomarker research in breast cancer to ensure clinical relevance. Nat Rev Cancer, 2007. 7(4): p. 309-15.

AIHW and NBCC, Breast cancer in Australia:
an overview, 2006. , in Cancer series no. 34.
cat. no. CAN 29. 2006, AIHW: Canberra.

Welfare, A.I.o.H.a., Cancer. 2000, Commonwealth Government Publications: Canberra.

McGuire, W.L. and G.M. Clark, Prognostic factors and treatment decisions in axillary-node-negative breast cancer. N Engl J Med, 1992. 326(26): p. 1756-61.

Talisuna, A.O., et al., Population-based validation of dihydrofolate reductase gene mutations for the prediction of sulfadoxine-pyrimethamine resistance in Uganda. Trans R Soc Trop Med Hyg, 2003. 97(3): p. 338-42.

Bloom, H.J. and W.W. Richardson, Histological grading and prognosis in breast cancer; a study of 1409 cases of which 359 have been followed for 15 years. Br J Cancer, 1957. 11(3): p. 359-77.
Elston, C.W. and I.O. Ellis, Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. Histopathology, 1991. 19(5): p. 403-10.
Burke, H.B. and D.E. Henson, Histologic grade as a prognostic factor in breast carcinoma. Cancer, 1997. 80(9): p. 1703-5; discussion 1706-7.
Nicholas, E.R., Histologic grade as a prognostic factor in breast carcinoma--reply. Cancer, 1997. 80(9): p. 1706-1707.
Roberti, N.E., The role of histologic grading in the prognosis of patients with carcinoma of the breast: is this a neglected opportunity? Cancer, 1997. 80(9): p. 1708-16.
Ellis, I.O., et al., Pathological prognostic factors in breast cancer. II. Histological type. Relationship with survival in a large study with long-term follow-up. Histopathology, 1992. 20(6): p. 479-89.
Frkovic-Grazio, S. and M. Bracko, Long term prognostic value of Nottingham histological grade and its components in early (pT1NOMO) breast carcinoma. J Clin Pathol, 2002. 55: p. 88-92.
Schumacher, M., et al., The prognostic effect of histological tumor grade in node-negative breast cancer patients. Breast Cancer Res Treat, 1993. 25(3): p. 235-45.
Lundin, J., et al., Omission of histologic grading from clinical decision making may result in overuse of adjuvant therapies in breast cancer: results from a nationwide study. J Clin Oncol, 2001. 19(1): p. 28-36.
Harvey, J.M., N.H. de Klerk, and G.F. Sterrett, Histological grading in breast cancer: interobserver agreement, and relation to other prognostic factors including ploidy. Pathology, 1992. 24(2): p. 63-8.
Robbins, P., et al., Histological grading of breast carcinomas: a study of interobserver agreement. Hum Pathol, 1995. 26(8): p. 873-9.
Frierson, H.F., Jr., et al., Interobserver reproducibility of the Nottingham modification of the Bloom and Richardson histologic grading scheme for infiltrating ductal carcinoma. Am J Clin Pathol, 1995. 103(2): p. 195-8.
Dalton, L.W., D.L. Page, and W.D. Dupont, Histologic grading of breast carcinoma. A reproducibility study. Cancer, 1994. 73(11): p. 2765-70.

Dalton, L.W., et al., Histologic grading of breast cancer: linkage of patient outcome with level of pathologist agreement. Mod Pathol, 2000. 13(7): p. 730-5.
Younes, M. and R. Laucirica, Lack of prognostic significance of histological grade in node-negative invasive breast carcinoma. Clin Cancer Res, 1997. 3(4): p. 601-4.
Hayes, D.F., C. Isaacs, and V. Stearns, Prognostic factors in breast cancer: current and new predictors of metastasis. J Mammary Gland Biol Neoplasia, 2001. 6(4): p. 375-92.
Sotiriou, C., et al., Gene Expression Profiling in Breast Cancer: Understanding the Molecular Basis of Histologic Grade To Improve Prognosis. J Natl Cancer Inst, 2006. 98(4): p. 262-272.
Ivshina, A.V., et al., Genetic Reclassification of Histologic Grade Delineates New Clinical Subtypes of Breast Cancer. Cancer Res, 2006. 66(21): p. 10292-10301.
Ma, X., et al., Gene expression profiles of human breast cancer progression. Proc Natl Acad Sci, 2003. 100(10): p. 5974-5979.
Filho, O.M., M. Ignatiadis, and C. Sotiriou, Genomic Grade Index: An important tool for assessing breast cancer tumor grade and prognosis. Crit Rev Oncol Hematol, 2010.
Ignatiadis, M. and C. Sotiriou, Understanding the molecular basis of histologic grade. Pathobiology, 2008. 75(2): p. 104-11.
Toussaint, J., et al., Improvement of the clinical applicability of the Genomic Grade Index through a qRT-PCR test performed on frozen and formalin-fixed paraffin-embedded tissues. BMC Genomics, 2009. 10: p. 424.
Ma, X.-J., et al., A Five-Gene Molecular Grade Index and HOXB13:IL17BR Are Complementary Prognostic Factors in Early Stage Breast Cancer. Clin Cancer Res, 2008. 14(9): p. 2601-2608.
Jerevall, P.L., et al., Prognostic utility of HOXB13:IL17BR and molecular grade index in early-stage breast cancer patients from the Stockholm trial. Br J Cancer, 2011. 104(11): p. 1762-9.

Chapter 8

PLANT, H. C., KASHYAP, A. S., MANTON, K. J., HOLLIER, B. G., HURST, C. P., STEIN, S. R., FRANCIS, G. D., BEADLE, G. F., UPTON, Z. & LEAVESLEY, D. I. 2014. Differential subcellular and extracellular localisations of proteins required for insulin-like growth factor- and extracellular matrix-induced signalling events in breast cancer progression. BMC Cancer, 14, 627.

8.1.1. Introduction

Epithelial-mesenchymal transition (EMT) has been associated with circulating tumour cells and metastasis in epithelial cancers including breast cancer. A number of biomarkers have been evaluated and loss of E-cadherin has been associated with lymph node metastases (Markiewicz et al., 2014). Markers of EMT were evaluated in this project to identify differences in expression and correlation with patient out come and clinicaopathological features.

8.1.2. Results

Abstract

Introduction: Cancer metastasis is the main contributor to breast cancer fatalities as women with the metastatic disease have poorer survival outcomes than women with localised breast cancers. There is an urgent need to develop appropriate prognostic methods to stratify patients based on the propensities of their cancers to metastasise. The insulin-like growth factor (IGF)-I:IGF binding protein (IGFBP):vitronectin complexes have been shown to stimulate changes in gene expression favouring increased breast cancer cell survival and a migratory phenotype. We therefore investigated the prognostic potential of these IGF- and extracellular matrix (ECM) interaction-induced proteins in the early identification of breast cancers with a propensity to metastasise using patient-derived tissue microarrays.

Methods: Semiquantitative immunohistochemistry analyses were performed to compare the extracellular and subcellular distribution of IGF- and ECM-induced signalling proteins among matched normal, primary cancer and metastatic cancer formalin-fixed paraffin-embedded breast tissue samples.

Results: The IGF- and ECM-induced signalling proteins were differentially expressed between subcellular and extracellular localisations. Vitronectin and IGFBP-5 immunoreactivity was lower while β1 integrin immunoreactivity was higher in the stroma surrounding metastatic cancer tissues, as compared to normal breast and primary cancer stromal tissues. Similarly, immunoreactive stratifin was found to be increased in the stroma of primary as well as metastatic breast tissues.

Immunoreactive fibronectin and β1 integrin was found to be highly expressed at the leading edge of tumours. Based on the immunoreactivity it was apparent that the cell

signalling proteins AKT1 and ERK1/2 shuffled from the nucleus to the cytoplasm with tumour progression.
Conclusion: This is the first in-depth, compartmentalised analysis of the distribution of IGF- and ECM-induced signalling proteins in metastatic breast cancers. This study has provided insights into the changing pattern of cellular localisation and expression of IGF- and ECM-induced signalling proteins in different stages of breast cancer. The differential distribution of these biomarkers could provide important prognostic and predictive indicators that may assist the clinical management of breast disease, namely in the early identification of cancers with a propensity to metastasise, and/or recur following adjuvant therapy.

Introduction

Experimental and clinical evidence has implicated a role for the insulin-like growth factor (IGF) axis in cancer progression [1]. In fact a number of inhibitors of, and antibodies directed against, the IGF type I receptor (IGF-IR) have been reported to show anti-tumour activity in vitro and in vivo, and are currently in clinical trials [2]. These studies, combined with many others, have highlighted the complexity of the dysregulation of the IGF system in cancers. Simply targeting the IGF-IR or the IGF system in isolation may therefore not be the most efficacious strategy for treating this disease; more complex therapeutic approaches to target the IGF system and prevent tumorigenesis and, in particular, metastasis, are likely to be required.
Cancer metastasis is the main contributor to breast cancer fatalities [3]. Women with metastatic breast cancers have considerably poorer survival outcomes than women whose cancers are localised to the breast [4, 5]. Adjuvant systemic therapies for patients with breast cancer metastasis remain palliative [3]. Understanding the processes underpinning the progression of breast cancer, identifying patients likely to develop metastases and developing strategies to prevent the secondary spread of cancers are of significant clinical and financial relevance. There has also been a growing urgency to create cost-effective and appropriate prognostic methods that can accurately resolve those patients with a poor prognosis that require more intense treatment regimes. The prognostic methods currently available are unable to adequately address this issue [6, 7].
A critical component that is often overlooked during the identification and analysis of prognostic biomarkers, and one which may explain the inability to develop adequate prognostic techniques thus far, is the interplay between tumour cells, the surrounding

microenvironment and the growth factors present in this milieu. Cellular attachment and interactions with the extracellular matrix (ECM) regulate biological responses vital for tumour progression. Considerable evidence indicates that interactions between proteins required for IGF-induced signalling events and those within the ECM contribute to processes leading to cancer progression. Studies by Kricker et al. [8] found that IGF-I stimulates migration of MCF-7 breast cancer cells when bound to the ECM protein vitronectin (VN) indirectly through the presence of IGF binding proteins (IGFBPs). The presence of function blocking antibodies against IGF-IR and VN-binding integrins abolished the enhanced migration of these cells [9], while, the overexpression of total-akt/protein kinase B (AKT) and phosphorylated-akt/protein kinase B (P-AKT) enhanced IGF-I:IGFBP:VN-stimulated migration [9]. Gene microarray technology has also been applied to elucidate the molecular mechanisms involved in IGF-I:IGFBP:VN-stimulated migration of breast cancer cells in in vitro cell based assays [10]. These studies have identified a number of genes, including Stratifin (SFN), enhancer-of-split and hairy-related protein 2 (Sharp-2), Tissue Factor, Claudin-1 (CLDN1), that are uniquely regulated by the IGF-I:IGFBP:VN complex.

The genes are known for their roles in migration, invasion as well as cell survival. However, to date the effects of IGF and ECM protein interactions on the dissemination and progression of breast cancer in vivo are unclear. Given this, we chose to investigate the clinical relevance of proteins required for IGF-induced signalling events and those within the ECM for the development and progression of breast cancer, as well as investigate these proteins as potential prognostic biomarkers.

Patients, Materials and Methods

Human ethics approval: Ethical approval for this work was obtained from the Queensland University of Technology, Australia (0800000565), the Princess Alexandra Hospital Australia (2005/163), Royal Brisbane & Women's Hospital, Australia (PR07/004) and Queensland Institute of Medical Research, Australia (P716). This project utilised archived human tissue samples collected between January 1970 and June 2005. The human tissue samples and patients records were collected as a routine part of clinical management of the breast disease. Patient consent was not required. All patient clinical information was obtained from the Queensland Cancer Registry (Australia) in a de-identified and encoded

manner. Approval to use these samples and data was sought from Dr Glenn Francis and Queensland Health (Australia).

Selection of patient specimen: This project utilised formalin-fixed paraffin-embedded (FFPE) archival breast carcinoma specimens from 91 women who presented with metastatic breast carcinoma (refer to Additional file 1 and 2 for further details).

These specimens were surgically removed from the breast and axillary lymph nodes (LNs). For each patient, tissues containing normal breast epithelial ducts, primary breast carcinoma and LN metastasis were identified from haematoxylin and eosin stained sections. Cores containing DCIS tissues were omitted due to low samples numbers. Details on the selected patient cohort are provided in Additional files 3 and 4.

Tissue microarray (TMA) construction: See details of TMA construction in Additional file 5. Where possible, the TMA cores were obtained from the leading edge of the tumour, thought to be where interactions between ligands in the ECM and the cancer cells were more likely to have functional significance [11].

Candidate biomarkers: The candidate molecules selected for this investigation were: IGFBP-5, VN, fibronectin (FN), αv integrin, β1 integrin, total-akt/protein kinase B 1 (Total-AKT1), P-AKT (Ser473), extracellular signal-related kinase-1 and extracellular signal-related kinase-2 (ERK1/2), phosphorylated-extracellular signal-related kinase-1 and extracellular signal-related kinase-2 (P-ERK1/2) (Thr202/Thr204), SHARP-2 and SFN. Oestrogen receptor (ER), progesterone receptor (PR) and HER2 were also selected for investigation.

Immunohistochemistry (IHC): The candidate markers were detected using commercially documented antibodies based upon prior independent validation for immunohistological applications in FFPE sections. Refer to Additional file 6 and Additional file 7 for specific details on the antibodies and IHC optimisation protocols, respectively.

Distiller: A secure, web based, flexible information management system: The virtual TMA slide files created using the NanoZoomer 2.0 series (Hamamatsu®, Hamamatsu City, Shizuoka Pref., Japan) digital slide scanner and scanning software NDP.scan 2.0 series (Hamamatsu®) were uploaded into Distiller (SlidePath Ltd Digital Pathology Solutions, Santry, Dublin, Ireland) for image analysis. Distiller was used to facilitate the integration of clinical records, research data, digital TMA slides and different data types into a hierarchical database (see Additional file 8 for information).

Scoring immunohistochemical immunoreactivity: The digital TMA images were examined and scored by trained anatomical pathology (AP) registrars without prior knowledge of the patient's clinical data (i.e. 'blind') within the Distiller framework. If there were no pathologists available to score the TMAs, they were scored by Helen C Plant. Qualitative differences in the immunoreactivity of the proteins within the cytoplasm, nucleus and membrane of the cells were determined for each TMA core containing either normal breast epithelial ducts (normal), primary breast carcinoma (primary) or metastatic breast carcinoma (LN met). Qualitative differences in staining of the stromal cells and ECM adjacent to normal, primary and LN met tissue were also recorded. Protein immunoreactivity was evaluated semiquantitatively using five scoring methods. These included: presence of protein immunoreactivity; intensity of protein immunoreactivity; percentage of cells with protein immunoreactivity; percentage class and quickscore (Q score) scoring method [12]. Details on these scoring methods and data consolidation strategies are listed in Additional file 9.

Statistical data analysis: PASW Statistics 18 version 18.0.2 (SPSS, IBM Corporation, Chicago, Illinois, USA) was used to evaluate statistical confidence of the data. The choice of test of association for the five scoring methods of protein immunoreactivity depended on the measurement scale of the scoring method. Presence is a binary outcome (present/absent) hence Pearson's $\chi 2$ test of independence was used. For the ordinal scaled intensity, a Kruskal-Wallis test was employed to determine if any of the groups demonstrated differences. No protected rank-based non-parametric test exists for the post-hoc evaluation of pair-wise differences. Instead, the Mann-Whitney U test was used to evaluate between-groups differences with inflation of family-wise type I error being controlled using Bonferroni corrections. Finally, the remaining three measures of protein immunoreactivity, percentage, percentage class and Q score were all treated as quantitative outcomes and one-way Analysis of Variance (ANOVA) followed by Tukey's HSDs for post-hoc testing was used to detect differences. As no rank-based non-parametric method exists to test for effect modification (i.e. interactions), interactions were probed by running the (one-way) Kruskal-Wallis tests for each strata of a potential effect modifier. For all tests, a significance level (α) of 0.05 was used, with the exception of where the Mann- Whitney U was used to test for post-hoc differences, where $\alpha FW = \alpha/k = 0.05/6 = 0.008$ was used ($k = 6$ represents the number of pairwise comparisons).

Results

The capability of the ECM and IGF system proteins to regulate cell function, and consequently tumorigenesis, is highly influenced by their spatial arrangement within and around the cell. It was observed that proteins required for IGF- and ECM-induced signalling events are differentially expressed between subcellular and extracellular localisations and that the interpretation of the protein immunoreactivity data is influenced by the scoring method applied. The results described below will only refer to the results obtained for the Q score scoring method [12]. The Q score values (x), including the standard deviation (SD) and sample numbers (n) for each protein across the tissue types and cellular localisation are outlined in Table 1.

Changes in ECM proteins: The most obvious differences in the immunoreactive distribution between normal breast, primary and metastatic cancer tissue samples was observed in proteins located in the extracellular space surrounding normal breast ducts and primary and metastatic tumours. These findings are intriguing given that the processes occurring during normal breast development are tightly regulated by the ECM and that the ability of the ECM to provide homeostatic regulation is disrupted during the development and progression of breast cancer. It was observed that the immunoreactivity of key ECM molecules, IGF regulators and integrins decreased with tumour development and/or progression. Significant differences in the immunoreactivity of stromal VN ($p < 0.001$), IGFBP-5 ($p < 0.001$) and β1 integrin ($p< 0.001$) within the tissue types examined were detected (Figure 1 A, 1 C and 2 Ai, respectively). Stromal IGFBP-5 and VN immunoreactivity in the metastatic cancer tissues was found to be significantly less than stromal IGFBP-5 and VN immunoreactivity in the normal breast tissues ($p < 0.001$ and $p < 0.001$, respectively) and primary cancer tissues ($p < 0.01$ and $p < 0.001$, respectively) (Figure 1 C and 1A, respectively). Additionally, stromal VN immunoreactivity was greater within normal breast tissue as compared to the immunoreactivity detected in primary cancer tissues ($p < 0.001$) (Figure 1 A). Despite not reaching statistical significance ($p = 0.054$), comparable trends were observed for stromal αv integrin staining with increasing invasiveness of the tissue types examined (Figure 1 B).

In contrast, the β1 integrin immunoreactivity detected in the stroma of metastatic cancer tissue was significantly higher than the β1 integrin immunoreactivity detected in the stroma in the normal breast ($p < 0.001$) and primary cancer ($p < 0.001$) tissue samples (Figure 2 Ai). In the primary cancer tissues, stromal reactivity of the β1 integrin was

significantly greater than within the normal breast tissues ($p < 0.001$) (Figure 2 Ai). No statistically significant differences in FN reactivity between the various tissue types were examined ($p = 0.094$) (Figure 2 Bi). These findings suggest that VN, IGFBP-5 and αv integrin reactivity in the stroma decreased while stromal β1 integrin immunoreactivity increased with tumour progression. Figure 2 Bi reveals trends, albeit not statistically significant, which suggest that the stromal localisation of FN differs to that of the other ECM proteins analysed.

Tumour leading edge: Given these findings, we next investigated whether the distribution of β1 integrin and FN immunoreactivity within the stroma could be functionally associated with cancer invasion. FN immunoreactivity was observed throughout the stroma immediately adjacent and distal to the leading edges of each tumour (Figure 2 Bii - v). There was also a higher presence of FN immunoreactivity both inside tumour cells at the leading edges and in the stroma directly surrounding the leading edges (Figure 2 Bii - v). In particular, greater membrane and cytoplasmic FN was associated with tumour cells at the leading edge and in close proximity to the leading edge, in contrast to the cells within the middle of the tumour. Paralleling the distribution of FN, β1 integrin immunoreactivity was detected throughout the stroma immediately surrounding and distant to the leading edges of the tumours (Figure 2 Aii - ix). There were many instances where the β1 integrin was also detected both inside tumour cells at the leading edges and within the stroma of the leading edges of tumours (Figure 2 Aii - ix). Again, greater membrane β1 integrin immunoreactivity was observed in tumour cells at the leading edge and in close proximity to the leading edge, compared to the main bulk of the tumour. However, there were no obvious differences between the cytoplasmic expression of β1 integrin in cells at the leading edge of tumours and those in the centre of the tumours.

SFN in stroma: Significant differences were evident in the immunoreactivity of SFN in the stroma of the tissue types examined ($p < 0.001$) (Figure 3). In particular, SFN immunoreactivity scores within the stroma of normal breast tissue were significantly lower than the SFN immunoreactivity scores within stroma of primary ($p < 0.05$) and metastatic cancer tissues ($p < 0.001$). In addition, this finding suggests that SFN immunoreactivity increases in the stroma with tumour development and progression.

Intracellular movement of cell signalling proteins: The occupation of IGF-IR and integrin molecules results in the recruitment of adapter proteins to the cell membrane and the

formation of multiprotein complexes and facilitates the phosphorylation and activation of signalling cascades including AKT and MAPK [13]. The phosphorylation of AKT and MAPK impacts the cellular localisation, specificity and consequently the protein targets of these signalling molecules [14, 15]. In light of this, the immunoreactivity of Total- and Phosphorylated- AKT and ERK1/2 within the cytoplasm and nucleus of cells from normal breast, primary and metastatic cancers were investigated to determine their role in the downstream signalling events during the development and progression of breast cancer. The intracellular localisation of SHARP-2 and SFN species uniquely regulated by the IGF-I:IGFBP:VN complex [10], were also investigated.

Significant differences in nuclear localisation of P-AKT, ERK1/2 and SHARP-2 reactivity ($p < 0.01$, $p < 0.05$ and $p < 0.001$, respectively) were detected between the normal breast, primary cancer and metastatic cancer tissues examined (Figure 4 C, D and 5 A). More specifically, nuclear P-AKT, ERK1/2 and SHARP-2 immunoreactivity within metastatic cancer tissues was lower than that observed within normal breast tissues ($p < 0.05$, $p < 0.05$ and $p < 0.01$, respectively) (Figure 4 C, D and 5 A). There was also less nuclear SHARP-2 immunoreactivity within metastatic breast tissue than within primary cancer tissues ($p < 0.01$). Thus, the nuclear localisation of P-AKT, ERK1/2 and SHARP-2 decreased with tumour development and/or progression. In contrast, Figures 4 and 5 reveal trends, albeit not statistically significant, which suggest that the cytoplasmic localisation of Total- AKT1, P-AKT and SHARP-2 may increase with tumour development and/or progression. Additionally, in many of the primary cancer and metastatic (data not shown) tissues in this study there was more nuclear immunoreactivity for SHARP-2 at the periphery of the tumour than was evident in the centre of the tumour (Figure 5 B).

We further observed that differences in nuclear and cytoplasmic SFN reactivity ($p <0.05$ and $p < 0.001$, respectively) between the normal breast, primary cancer and metastatic cancer tissues were significant (Figure 4 A). In particular, nuclear and cytoplasmic SFN within normal breast tissue was lower than the amount of nuclear and cytoplasmic SFN within primary cancer ($p < 0.05$ and $p < 0.001$, respectively) and metastatic cancer tissues ($p < 0.01$ and $p < 0.001$, respectively). There were no significant differences in nuclear and cytoplasmic immunoreactivity of SFN between primary and metastatic tumours.

Internalisation of ECM proteins: Our data also provides evidence that ECM molecules are internalised during breast cancer development and metastasis. Statistically significant differences in cytoplasmic ($p < 0.05$) VN immunoreactivity was observed between the specific tissue types examined in this study (Figure 6). In particular, lower cytoplasmic VN immunoreactivity was observed in normal breast tissues than cytoplasmic VN immunoreactivity within metastatic cancer tissues ($p < 0.05$) (Figure 6). As such, our data indicates that VN redistributes to the cytoplasm with tumour progression.

Discussion

In this study a change in the immunoreactivity of key ECM molecules, IGF regulators and integrins was observed with breast tumour progression. These observations suggest that the ECM surrounding normal breast ductal structures is remodelled during tumour development and progression. The ECM protein FN and the cell surface β1 integrin (the FN-binding receptor) were found to be highly expressed along the leading edge of many primary tumours in the present study. Interestingly, FN is implicated in epithelial-to-mesenchymal transition (EMT) [16] and both the β1 integrin and FN are required in the formation of lamellipodia, filopodia and invadopodia [17, 18]; potentially supporting their role in ECM remodelling and subsequent tumour cell invasion [19]. Indeed, the β1 integrin has been reported to be more highly expressed in primary tumours with LN metastases [20] and with poor survival outcomes [21]. This fits in well with our findings that suggest an increase in β1 integrin immunoreactivity with increasing invasiveness. Various proteases, including matrix metalloproteinases (MMPs) have been implicated in ECM remodelling events that allow cancer cells to migrate [22]. It has been shown that the expression, activity and/or internalisation of MMPs is regulated by integrin-ECM interactions in endothelial cells [23]. Integrins, such as the αvβ3 integrin, cooperate with MMPs to regulate breast cancer cell migration [24]. Interestingly, IGF- I:VN:IGFBP-5-stimulated breast cell migration, which requires the IGF-IR and VN- binding integrins [9], can regulate the gene expression of proteases such as MMP13, MMP7, ADAMTS5, CPM and protease inhibitors such as SERPINE1 and TRP1 ([10] and supplementary data from [10]).

The data we report here has also provided evidence that ECM molecules and their associated membrane-bound receptors, the integrins, are internalised during breast cancer development and metastasis. Step-wise increases in cytoplasmic and concomitant decreases in stromal immunoreactivity of VN and the αv integrin (data not shown) were

evident between normal breast, primary and metastatic cancer tissues. As described previously, VN can be internalised by integrin receptor- mediated endocytosis and degraded within the lysosomes [25]. There is also evidence indicating that the VN-binding αv integrin can be recycled to the cell membrane through intracellular signals [26]. This decrease in the stromal VN and the concomitant increase in the cytoplasmic VN with breast cancer progression suggests a potential re-shuffling or trafficking of VN from the tumour stroma to the tumour cell cytoplasm with increasing invasiveness of the tumour-type.

Given the importance of the phosphoinositol 3-kinase (PI3K) and mitogen-activated protein kinase (MAPK) signalling pathways in IGF-I:IGFBP:VN-stimulated migration of breast cancer cells in vitro [9], and in IGF-I-stimulated ECM re- modelling [27] and EMT events [27, 28], protein intermediates within these pathways were also investigated. Decreases in the nuclear immunoreactivity and increases in the cytoplasmic immunoreactivity in breast cancer tissues were observed in this study.

We propose a number of explanations for these findings: namely, the preferential activation of substrates in the cytoplasm (of cancer cells) rather than in the nucleus; enzyme-mediated dephosphorylation; and protein internalisation. Both AKT1 and MAPK contain transportation signals which potentially enable their movement throughout a cell [29, 30]. Defects in these transportation signals during cancer tumorigenesis might explain the results of this study. Ras homolog gene family member B (RhoB), which has been shown to influence the trafficking of Total- and P- AKT in primary human endothelial cells [31], may impede the import of Total- and P- AKT into or promote the export of Total- and P-AKT from the nucleus of breast cancer cells; resulting in an accumulation of AKT in cytoplasmic compartments.

Phospho-kinases, such as MAPK, are not necessarily required to enter the nucleus to regulate gene transcription. In fact, the activation of transcription factors in the cytoplasm and their movement into the nucleus for transcriptional control [32] has been reported. The duration and strength of AKT and MAPK signalling in breast epithelial cells can also be regulated in different subcellular locations through the action of various cytoplasmic and nuclear phosphatases [33]. Indeed, phosphatase and tensin homolog (PTEN), a dual lipid and protein phosphatase, can be localised to the cell nucleus [34], and if functional, may therefore de-phosphorylate nuclear AKT. As discussed by Tzivion et al. [35], the ability of a protein to interact with modifying enzymes, such as phosphatases, can be influenced by the presence of 14-3-3 proteins.

Intriguingly, 14-3-3 proteins, including SFN, have been shown to regulate the cytoplasmic sequestration and nuclear retention of cell cycle regulators, many of which are associated with and downstream of the PI3K pathway [36]. It is highly likely that 14-3-3 proteins, such as SFN, may regulate similar sequestration events for P-AKT itself. It was intriguing to find that the increases in the nuclear and cytoplasmic immunoreactivity of SFN with breast cancer development and progression as measured in this study correlated with the increase in mRNA expression of SFN reported by Kashyap et al. [10]. In contrast, other studies have reported the downregulation of SFN expression in breast cancers [37, 38].

However, Neal et al. [39] showed that overexpression of SFN reduces the overall and disease- free survival of breast cancer patients and is able to predict which patients have a high susceptibility to develop metastasis. Interestingly, it has been proposed that 14-3-3 proteins are important regulators of external environmental signals by eliciting positive and negative effects on the IGF signalling pathway. The ability of 14-3-3 proteins to bind phospho-serine enables them to bind to the IGF-IR [40]. Yang et al. [41] have shown that SFN can also bind to and inhibit the activity of AKT, preventing AKT-mediated cellular events. They also indicate that SFN expression was inversely correlated with P-AKT expression [41]; this supports our findings of decreases in nuclear P-AKT and increases in intracellular SFN with tumour development and progression.

We were intrigued to find that SFN was differentially expressed within the stroma surrounding the tissue types examined. Although previous reports of SFN expression have been limited to the cytoplasm of malignant breast cells [38], in vitro evidence indicates that SFN can be excreted by keratinocytes into the pericellular matrix [42]. Extracellular SFN is a key regulator of MMP function and ECM degradation. Studies have shown that following the release of SFN from keratinocytes, MMP-1 mRNA [43] and MMP-1 protein synthesis [44] increases in dermal fibroblasts. Increases in mRNA encoding the β1 integrin have also been observed in dermal fibroblasts after treatment with SFN, or in co-cultures with keratinocytes known to release SFN [45]. Under the same conditions the expression of many ECM molecules, including collagen type I, FN and α1 integrin, decreases. This collective evidence suggests that it is highly likely that SFN may mediate similar functions regulating the degradation of ECM during epithelial tumour development and progression.

We also observed step-wise increases in cytoplasmic and decreases in nuclear immunoreactivity of SHARP-2 between the normal breast, primary and metastatic cancer tissues; this may be explained by nucleocytoplasmic shuttling events. SHARP- 2 is known to possess a functional nuclear export sequence (NES) and two nuclear localisation signal (NLS) motifs [46]. A study by Ivanova et al. [46] suggests that SHARP-2 may be required in the nucleus of proliferating and differentiating cells to regulate gene transcription after stimulation by an external factor. They also propose that SHARP-2 may be sequestered in the cytoplasm following cell differentiation.

In summary, we have reported changes in the temporal and spatial distribution of IGF- and ECM-induced signalling proteins that occur during breast cancer metastasis. Specifically, our findings provide further evidence that the ECM surrounding normal breast ductal

structures is remodelled during tumour development and progression, and that FN and the β1 integrin are important for the formation of invadopodia and for the epithelial-to-mesenchymal transition events (shown by others [47]) to support dissemination. Analysis of stromal and subcellular SFN immunoreactivity suggested a causal relationship in ECM remodelling events and the localisation and activity of proteins important for IGF- and ECM-induced signalling cascades. It also appears plausible that in cells at the leading edge of tumours, SHARP-2 moves into the nucleus to repress the transcription of genes associated with the hypoxic response.

Collectively the above data highlight the possibility that there are broader biological implications of, and explanations for, the differential immunoreactivity of IGF signalling and ECM components in the stroma and/or in subcellular locations within normal breast, primary breast cancer and metastatic breast cancers. This is highly pertinent given that protein function and protein localisation are closely correlated. Studies have also shown that accounting for protein localisation can be an essential requirement to identifying correlations with other proteins when applying the IHC technique [48]. To date, very few studies have evaluated the prognostic significance of differential protein distribution within diagnostic breast cancer tissue samples.

Early studies do, however, suggest that specific locations of specific proteins associated with the IGF signalling cascade and the ECM have shown potential as markers for patient prognosis and therapeutic response [49, 50]. We argue that to date the potential of many molecular species to serve as markers of patient prognosis and therapeutic response is being missed by overlooking their subcellular/extracellular distribution. In view of this, we recommend a more complete analysis of protein localisation within diagnostic pathology and improvements to reporting and inclusion of protein localisation in routine pathological examinations as our data indicates the potential role of protein localisation in the progression of disease.

Conclusions

There is potential that the cellular and ECM events outlined herein could be manipulated to provide clinical benefits and improve the clinical management of breast cancer. In particular, may lead to early prognostic and predictive identification of patients with poor survival outcomes. However, prior to this occurring, the prognostic significance of the cellular and ECM events reported in this study must be identified.

Competing interests

ZU, DIL and BGH and have purchased shares in Tissue Therapies Ltd. (TIS), an enterprise spun-out from the Queensland University of Technology, Brisbane, to commercialise some of the technology described in this manuscript. ZU and DIL consult for TIS and are named inventors on patents licensed to TIS. GDF has shares in Prognostic Pathology Research Facility and Bioprognostics P/L, private companies with pre-existing IP in development of the database. HCP, ASK, KJM, CPH, SRS, and GFB have nothing to declare.

Authors' contributions

HCP carried out the experiments and performed data collection, analysis, and interpretation. ASK was involved in the drafting and editing of the manuscript and contributed to interpretation of the data. KJM and BGH contributed to the experimental design and interpretation of the data. CPH designed and performed statistical analyses of the data and contributed to the interpretation of the same. SRS contributed to the experimental design, assisted with data collection and analysis.
GDF and GFB sourced the experimental material and contributed to the collection of data, its analysis and interpretation. ZU contributed to the experimental design and interpretation of the data. DIL contributed to the experimental design, the interpretation and analysis of the data and co-ordinated this project. All authors provided editorial assistance and have read and approved this manuscript.

Acknowledgements

The authors would like to acknowledge Sarah Wagner from the Molecular and Clinical Pathology Research Laboratory (MaCH R), Health Services Support Agency (HSSA), Queensland Health based at the Princess Alexandra Hospital, Brisbane for laboratory assistance. This project was supported by Wesley Research Institute Grants (#2007-06 and #2010-04), Australia.

References

1. Samani AA, Yakar S, LeRoith D, Brodt P: The role of the IGF system in cancer growth and metastasis: overview and recent insights. Endocr Rev 2007, 28(1):20-47.
2. Yuen JS, Macaulay VM: Targeting the type 1 insulin-like growth factor receptor as a treatment for cancer. Expert Opin Ther Targets 2008, 12(5):589-603.
3. DeVita V, Hellman S, Rosenburg S (eds.): Cancer: Principles & Practice of Oncology, 7th Edition edn. Philadelphia: Lippincott Williams & Wilkins; 2005.
4. AIHW: Breast cancer survival by size and nodal status in Australia. In. Edited by Centre AIoHaWNBC, vol. Cancer Series no. 39. Canberra: AIHW; 2007.
5. Howlader N, Noone AM, Krapcho M, Neyman N, Aminou R, Waldron W, Altekruse SF, Kosary CL, Ruhl J, Tatalovich Z et al (eds.): SEER Cancer Statistics Review, 1975-2008. Bethesda, MD: National Cancer Institute; 2011.
6. Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, Baehner FL, Walker MG, Watson D, Park T et al: A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med 2004, 351(27):2817-2826.
7. Glas AM, Floore A, Delahaye LJ, Witteveen AT, Pover RC, Bakx N, Lahti- Domenici JS, Bruinsma TJ, Warmoes MO, Bernards R et al: Converting a breast cancer microarray signature into a high-throughput diagnostic test. BMC Genomics 2006, 7:278.
8. Kricker JA, Towne CL, Firth SM, Herington AC, Upton Z: Structural and functional evidence for the interaction of insulin-like growth factors (IGFs) and IGF binding proteins with vitronectin. Endocrinology 2003, 144(7):2807-2815.
9. Hollier BG, Kricker JA, Van Lonkhuyzen DR, Leavesley DI, Upton Z: Substrate-bound insulin-like growth factor (IGF)-I-IGF binding protein- vitronectin-stimulated breast cell migration is enhanced by coactivation of the phosphatidylinositide 3-Kinase/AKT pathway by alphav-integrins and the IGF-I receptor. Endocrinology 2008, 149(3):1075-1090.
10. Kashyap AS, Hollier BG, Manton KJ, Satyamoorthy K, Leavesley DI, Upton Z: Insulin-like growth factor-I:vitronectin complex-induced changes in gene expression effect breast cell survival and migration. Endocrinology 2011, 152(4):1388-1401.
11. Gladson CL, Cheresh DA: Glioblastoma expression of vitronectin and the alpha v beta 3 integrin. Adhesion mechanism for transformed glial cells. J Clin Invest 1991, 88(6):1924-1932.

12. Detre S, Saclani Jotti G, Dowsett M: A "quickscore" method for immunohistochemical semiquantitation: validation for oestrogen receptor in breast carcinomas. J Clin Pathol 1995, 48(9):876-878.
13. Ivaska J, Heino J: Cooperation between integrins and growth factor receptors in signaling and endocytosis. Annu Rev Cell Dev Biol 2011, 27:291-320.
14. Rosner M, Hanneder M, Freilinger A, HengstschlÃ¤ger M: Nuclear/cytoplasmic localization of Akt activity in the cell cycle. Amino Acids 2007, 32(3):341-345.
15. Roux PP, Blenis J: ERK and p38 MAPK-Activated Protein Kinases: a Family of Protein Kinases with Diverse Biological Functions. Microbiology and Molecular Biology Reviews 2004, 68(2):320-344.
16. Zeisberg M, Neilson EG: Biomarkers for epithelial-mesenchymal transitions. J Clin Invest 2009, 119(6):1429-1437.
17. Destaing O, Planus E, Bouvard D, Oddou C, Badowski C, Bossy V, Raducanu A, Fourcade B, Albiges-Rizo C, Block MR: beta1A integrin is a master regulator of invadosome organization and function. Mol Biol Cell 2010, 21(23):4108-4119.
18. Mueller SC, Chen WT: Cellular invasion into matrix beads: localization of beta 1 integrins and fibronectin to the invadopodia. J Cell Sci 1991, 99 (Pt 2):213-225.
19. Nakahara H, Mueller SC, Nomizu M, Yamada Y, Yeh Y, Chen WT: Activation of beta1 integrin signaling stimulates tyrosine phosphorylation of p190RhoGAP and membrane-protrusive activities at invadopodia. J Biol Chem 1998, 273(1):9-12.
20. Leth-Larsen R, Lund R, Hansen HV, Laenkholm AV, Tarin D, Jensen ON, Ditzel HJ: Metastasis-related plasma membrane proteins of human breast cancer cells identified by comparative quantitative mass spectrometry. Mol Cell Proteomics 2009, 8(6):1436-1449.
21. Ioachim E, Charchanti A, Briasoulis E, Karavasilis V, Tsanou H, Arvanitis DL, Agnantis NJ, Pavlidis N: Immunohistochemical expression of extracellular matrix components tenascin, fibronectin, collagen type IV and laminin in breast cancer: their prognostic value and role in tumour invasion and progression. Eur J Cancer 2002, 38(18):2362-2370.
22. Kessenbrock K, Plaks V, Werb Z: Matrix metalloproteinases: regulators of the tumor microenvironment. Cell 2010, 141(1):52-67.
23. Galvez BG, Matias-Roman S, Yanez-Mo M, Sanchez-Madrid F, Arroyo AG: ECM regulates MT1-MMP localization with beta1 or alphavbeta3 integrins at distinct cell

compartments modulating its internalization and activity on human endothelial cells. J Cell Biol 2002, 159(3):509-521.
24. Rolli M, Fransvea E, Pilch J, Saven A, Felding-Habermann B: Activated integrin alphavbeta3 cooperates with metalloproteinase MMP-9 in regulating migration of metastatic breast cancer cells. Proc Natl Acad Sci U S A 2003, 100(16):9482-9487.
25. Panetti TS, McKeown-Longo PJ: The alpha v beta 5 integrin receptor regulates receptor-mediated endocytosis of vitronectin. J Biol Chem 1993, 268(16):11492-11495.
26. Roberts M, Barry S, Woods A, van der Sluijs P, Norman J: PDGF-regulated rab4-dependent recycling of alphavbeta3 integrin from early endosomes is necessary for cell adhesion and spreading. Curr Biol 2001, 11(18):1392- 1402.
27. Walsh LA, Damjanovski S: IGF-1 increases invasive potential of MCF 7 breast cancer cells and induces activation of latent TGF-beta1 resulting in epithelial to mesenchymal transition. Cell Commun Signal 2011, 9(1):10.
28. Sivakumar R, Koga H, Selvendiran K, Maeyama M, Ueno T, Sata M: Autocrine loop for IGF-I receptor signaling in SLUG-mediated epithelial- mesenchymal transition. Int J Oncol 2009, 34(2):329-338.
29. Saji M, Vasko V, Kada F, Allbritton EH, Burman KD, Ringel MD: Akt1 contains a functional leucine-rich nuclear export sequence. Biochem Biophys Res Commun 2005, 332(1):167-173.
30. Chuderland D, Konson A, Seger R: Identification and characterization of a general nuclear translocation signal in signaling proteins. Mol Cell 2008, 31(6):850-861.
31. Adini I, Rabinovitz I, Sun JF, Prendergast GC, Benjamin LE: RhoB controls Akt trafficking and stage-specific survival of endothelial cells during vascular development. Genes Dev 2003, 17(21):2721-2732.
32. Macfarlane WM, Smith SB, James RF, Clifton AD, Doza YN, Cohen P, Docherty K: The p38/reactivating kinase mitogen-activated protein kinase cascade mediates the activation of the transcription factor insulin upstream factor 1 and insulin gene transcription by high glucose in pancreatic beta-cells. J Biol Chem 1997, 272(33):20936-20944.
33. Bermudez O, Pages G, Gimond C: The dual-specificity MAP kinase phosphatases: critical roles in development and cancer. Am J Physiol Cell Physiol 2010, 299(2):C189-202.
34. Baker SJ: PTEN enters the nuclear age. Cell 2007, 128(1):25-28.

35. Tzivion G, Gupta VS, Kaplun L, Balan V: 14-3-3 proteins as potential oncogenes. Semin Cancer Biol 2006, 16(3):203-213.
36. Hermeking H, Benzinger A: 14-3-3 proteins in cell cycle regulation. Semin Cancer Biol 2006, 16(3):183-192.
37. Mirza S, Sharma G, Parshad R, Srivastava A, Gupta SD, Ralhan R: Clinical significance of Stratifin, ERalpha and PR promoter methylation in tumor and serum DNA in Indian breast cancer patients. Clin Biochem 2010, 43(4-5):380-386.
38. Simooka H, Oyama T, Sano T, Horiguchi J, Nakajima T: Immunohistochemical analysis of 14-3-3 sigma and related proteins in hyperplastic and neoplastic breast lesions, with particular reference to early carcinogenesis. Pathol Int 2004, 54(8):595-602.
39. Neal CL, Yao J, Yang W, Zhou X, Nguyen NT, Lu J, Danes CG, Guo H, Lan KH, Ensor J et al: 14-3-3zeta overexpression defines high risk for breast cancer recurrence and promotes cancer cell survival. Cancer Res 2009, 69(8):3425-3432.
40. Craparo A, Freund R, Gustafson TA: 14-3-3 (epsilon) interacts with the insulin-like growth factor I receptor and insulin receptor substrate I in a phosphoserine-dependent manner. J Biol Chem 1997, 272(17):11663- 11669.
41. Yang H, Wen YY, Zhao R, Lin YL, Fournier K, Yang HY, Qiu Y, Diaz J, Laronga C, Lee MH: DNA damage-induced protein 14-3-3 sigma inhibits protein kinase B/Akt activation and suppresses Akt-activated cancer. Cancer Res 2006, 66(6):3096-3105.
42. Chavez-Munoz C, Morse J, Kilani R, Ghahary A: Primary human keratinocytes externalize stratifin protein via exosomes. J Cell Biochem 2008, 104(6):2165-2173.
43. Ghahary A, Karimi-Busheri F, Marcoux Y, Li Y, Tredget EE, Taghi Kilani R, Li L, Zheng J, Karami A, Keller BO et al: Keratinocyte-releasable stratifin functions as a potent collagenase-stimulating factor in fibroblasts. J Invest Dermatol 2004, 122(5):1188-1197.
44. Lam E, Kilani RT, Li Y, Tredget EE, Ghahary A: Stratifin-induced matrix metalloproteinase-1 in fibroblast is mediated by c-fos and p38 mitogen- activated protein kinase activation. J Invest Dermatol 2005, 125(2):230- 238.
45. Ghaffari A, Li Y, Karami A, Ghaffari M, Tredget EE, Ghahary A: Fibroblast extracellular matrix gene expression in response to keratinocyte- releasable stratifin. J Cell Biochem 2006, 98(2):383-393.

46. Ivanova A, Liao SY, Lerman MI, Ivanov S, Stanbridge EJ: STRA13 expression and subcellular localisation in normal and tumour tissues: implications for use as a diagnostic and differentiation marker. J Med Genet 2005, 42(7):565-576.
47. Park J, Schwarzbauer JE: Mammary epithelial cell interactions with fibronectin stimulate epithelial-mesenchymal transition. Oncogene 2013.
48. Petricevic B, Vrbanec D, Jakic-Razumovic J, Brcic I, Rabic D, Badovinac T, Ozimec E, Bali V: Expression of Toll-like receptor 4 and beta 1 integrin in breast cancer. Med Oncol 2011.
49. Aleksic T, Chitnis MM, Perestenko OV, Gao S, Thomas PH, Turner GD, Protheroe AS, Howarth M, Macaulay VM: Type 1 insulin-like growth factor receptor translocates to the nucleus of human tumor cells. Cancer Res 2010, 70(16):6412-6419.
50. Yamashita H, Takahashi S, Ito Y, Yamashita T, Ando Y, Toyama T, Sugiura H, Yoshimoto N, Kobayashi S, Fujii Y et al: Predictors of response to exemestane as primary endocrine therapy in estrogen receptor-positive breast cancer. Cancer Sci 2009, 100(11):2028-2033.
51. Pilewski JM, Liu L, Henry AC, Knauer AV, Feghali-Bostwick CA: Insulin- like growth factor binding proteins 3 and 5 are overexpressed in idiopathic pulmonary fibrosis and contribute to extracellular matrix deposition. Am J Pathol 2005, 166(2):399-407.
52. Yamaguchi Y, Yasuoka H, Stolz DB, Feghali-Bostwick CA: Decreased caveolin-1 levels contribute to fibrosis and deposition of extracellular IGFBP-5. J Cell Mol Med 2011, 15(4):957-969.

Figure captions

Figure 1: Stromal immunoreactivity of VN, αv integrin and IGFBP-5. Immunoreactivity of VN (A), αv integrin (B) and IGFBP-5 (C) within the stroma surrounding normal breast (Normal), primary cancer (Primary) and LN metastasis tissues is depicted. Immunoreactivity was evaluated semiquantitatively using the Q score (intensity x percentage class, score: 0 – 18) method. Intensity of reactivity (score: 0 = negative; 1 = weak; 2 = moderate, and; 3 = strong). Percentage class (score: 1 = 0-4%; 2 = 5-19%; 3 = 20-39%; 4 = 40-59%; 5 = 60-79%; 6 = 80-100%).
Data are displayed using the mean ± 2 standard error (SE). Asterisks (** and ***) indicate statistically significant differences at $p<0.01$ and <0.001, respectively. Scale bar = 30 μm.

Figure 2: Stromal immunoreactivity of β1 integrin and FN. Immunoreactivity of β1 integrin is depicted in A. i) Immunoreactivity of β1 integrin within the stroma surrounding normal breast (Normal), primary cancer (Primary) and LN metastasis tissues. ii – ix) Representative images demonstrating distribution of β1 integrin in the stroma and/or cells at the leading edges of Normal (ii and iii), ductal carcinoma in situ (iv and v), Primary(vi and vii) and LN metastasis (viii and ix) tissues.
Immunoreactivity of FN is depicted in B. i) Immunoreactivity of FN within the stroma surrounding Normal, Primary and LN metastasis tissues. ii – v) Representative images demonstrating distribution of β1 integrin immunoreactivity in the stroma and/or cells at the leading edges of Primary (ii and iii) and LN metastasis (iv and v) tissues. Immunoreactivity was evaluated semiquantitatively using the Q score (intensity x percentage class, score: 0 – 18) method. Intensity of reactivity (score: 0 = negative; 1 = weak; 2 = moderate, and; 3 = strong). Percentage class (score: 1 = 0- 4%; 2 = 5-19%; 3 = 20-39%; 4 = 40-59%; 5 = 60-79%; 6 = 80-100%). Data are displayed using the mean ± 2 standard error (SE). Asterisks (* and ***) indicate statistically significant differences at $p<0.05$ and <0.001, respectively. The scale bar for Aii, Aiv, Avi, Aviii, Bii and Biv is 200 μm and for Aiii, Av, Avii, Aix, Biii and Bv is 30 μm.

Figure 3: Stromal immunoreactivity of SFN. Immunoreactivity of SFN within the stroma surrounding normal breast (Normal), primary cancer (Primary) and LN metastasis tissues is depicted. Immunoreactivity was evaluated semiquantitatively using the Q score (intensity x percentage class, score: 0 – 18) method. Intensity of reactivity (score: 0 = negative; 1 = weak; 2 = moderate, and; 3 = strong). Percentage class (score: 1 = 0-4%; 2 = 5-19%; 3 = 20-39%; 4 = 40-59%; 5 = 60-79%; 6 = 80-100%). Data are displayed using the mean ± 2 standard error (SE). Asterisks (* and ***) indicate statistically significant differences at $p<0.05$ and <0.001, respectively. Scale bar = 30 μm.

Figure 4: Tissue localisation. Differential localisation of SFN (A), Total-AKT1 (B), P-AKT (C), ERK1/2 (D) and P-ERK1/2 (E) is depicted. Nuclear (open circles) and cytoplasmic (closed triangles) immunoreactivity within normal breast (Normal), primary cancer (Primary) and LN metastasis tissues was determined. Antibody immunoreactivity was evaluated semiquantitatively using the Q score (intensity x percentage class, score: 0 – 18) method. Intensity of reactivity (score: 0 = negative; 1= weak; 2 = moderate, and; 3 =

strong). Percentage class (score: 1 = 0-4%; 2 = 5-19%; 3 = 20-39%; 4 = 40-59%; 5 = 60-79%; 6 = 80-100%). Data are displayed using the mean ± 2 standard error (SE). Asterisks (*, ** and ***) indicate statistically significant differences at p<0.05, <0.01 and <0.001, respectively.

Figure 5: SHARP-2 immunoreactivity. Immunoreactivity of SHARP-2 within the nucleus (open circles) and cytoplasm (closed triangles) of cells from normal breast (Normal), primary cancer (Primary) and LN metastasis tissues is depicted in A. Immunoreactivity was evaluated semiquantitatively using the Q score (intensity x percentage class, score: 0 – 18) method. Intensity of reactivity (score: 0 = negative; 1= weak; 2 = moderate, and; 3 = strong). Percentage class (score: 1 = 0-4%; 2 = 5-19%; 3 = 20-39%; 4 = 40-59%; 5 = 60-79%; 6 = 80-100%). Data are displayed using the mean ± 2 standard error (SE). Asterisks (**) indicate statistically significant differences at p<0.01. Representative images demonstrating the distribution of SHARP-2 in the cancer cells of primary cancer tissue samples is depicted in B (i-iv). The scale bar for A, Bii and Biv is 30 μm and for Bi and Biii is 200 μm.

Figure 6: VN immunoreactivity. Immunoreactivity of VN within the cell cytoplasm (open circles) and the stroma (closed triangles) of normal breast (Normal), primary cancer (Primary) and LN metastasis tissues is depicted. Immunoreactivity was evaluated semiquantitatively using the Q score (intensity x percentage class, score: 0– 18) method. Intensity of reactivity (score: 0 = negative; 1 = weak; 2 = moderate, and; 3 = strong). Percentage class (score: 1 = 0-4%; 2 = 5-19%; 3 = 20-39%; 4 = 40-59%; 5 = 60-79%; 6 = 80-100%). Data are displayed using the mean ± 2 standard error (SE). Asterisks (* and ***) indicate statistically significant differences at p<0.05 and <0.001, respectively. Scale bar = 30 μm.

Tables

Table 1: Q-score values for immunoreactivity of each protein across tissue types and cellular localisation

Protein	**Cellular localisation**	**Type of breast tissue**	***x***	***SD***	***n***
α_v integrin	Stroma	Normal	1.29	1.25	7
α_v integrin	Stroma	Primary	0.34	0.83	32

α_v integrin	Stroma	LN Metastasis	0.16	0.69	19
VN	Cytoplasm	Normal	0.78	1.65	23
VN	Cytoplasm	Primary	2.21	2.69	86
VN	Cytoplasm	LN Metastasis	2.44	2.38	68
VN	Stroma	Normal	7.13	3.75	23
VN	Stroma	Primary	2.84	2.73	86
VN	Stroma	LN Metastasis	0.84	2.13	68
IGFBP-5	Stroma	Normal	13.00	4.52	6
IGFBP-5	Stroma	Primary	8.49	4.18	35
IGFBP-5	Stroma	LN Metastasis	4.17	4.57	23
β_1 integrin	Stroma	Normal	4.79	3.62	14
β_1 integrin	Stroma	Primary	9.71	4.36	84
β_1 integrin	Stroma	LN Metastasis	14.47	4.21	68
FN	Stroma	Normal	4.08	1.88	12
FN	Stroma	Primary	8.66	4.42	85
FN	Stroma	LN Metastasis	7.36	5.05	67
SFN	Nucleus	Normal	1.07	1.73	14
SFN	Nucleus	Primary	3.40	3.27	82
SFN	Nucleus	LN Metastasis	3.88	2.84	68
SFN	Cytoplasm	Normal	2.93	2.30	14
SFN	Cytoplasm	Primary	5.43	2.20	82
SFN	Cytoplasm	LN Metastasis	5.75	1.93	68
SFN	Stroma	Normal	0.00	0.00	14
SFN	Stroma	Primary	0.78	0.89	82
SFN	Stroma	LN Metastasis	1.12	0.95	68
SHARP-2	Nucleus	Normal	8.29	3.71	14
SHARP-2	Nucleus	Primary	5.76	5.61	86
SHARP-2	Nucleus	LN Metastasis	3.33	3.69	70
SHARP-2	Cytoplasm	Normal	6.36	4.50	14
SHARP-2	Cytoplasm	Primary	7.90	3.72	86

SHARP-2	Cytoplasm	LN Metastasis	8.80	4.57	70
T-AKT1	Nucleus	Normal	5.69	6.91	13
T-AKT1	Nucleus	Primary	3.39	4.08	85
T-AKT1	Nucleus	LN Metastasis	2.59	2.93	68
T-AKT1	Cytoplasm	Normal	7.54	3.76	13
T-AKT1	Cytoplasm	Primary	8.78	3.95	85
T-AKT1	Cytoplasm	LN Metastasis	9.37	3.67	68
P-AKT	Nucleus	Normal	13.50	6.21	16
P-AKT	Nucleus	Primary	11.51	5.52	86
P-AKT	Nucleus	LN Metastasis	9.32	5.44	68
P-AKT	Cytoplasm	Normal	1.19	1.47	16
P-AKT	Cytoplasm	Primary	2.19	2.12	86
P-AKT	Cytoplasm	LN Metastasis	2.00	2.32	68
ERK1/2	Nucleus	Normal	2.67	2.92	15
ERK1/2	Nucleus	Primary	1.77	2.76	84
ERK1/2	Nucleus	LN Metastasis	0.75	1.82	67
ERK1/2	Cytoplasm	Normal	3.73	3.31	15
ERK1/2	Cytoplasm	Primary	6.18	4.20	84
ERK1/2	Cytoplasm	LN Metastasis	5.42	3.73	67
P-ERK1/2	Nucleus	Normal	1.83	2.79	12
P-ERK1/2	Nucleus	Primary	1.89	3.47	76
P-ERK1/2	Nucleus	LN Metastasis	0.41	1.03	64
P-ERK1/2	Cytoplasm	Normal	9.67	4.46	12
P-ERK1/2	Cytoplasm	Primary	10.08	4.27	76
P-ERK1/2	Cytoplasm	LN Metastasis	9.78	3.95	64

x = *Q score*; *SD* = standard deviation; *n* = sample size; LN = lymph node; P = phosphorylated; T = Total

Additional material information

Additional file 1 Format: PDF

Title and description: The actual number of normal breast epithelial duct, ductal carcinoma in situ (DCIS), primary breast carcinoma and/or lymph node (LN) metastasis tissues (n).

Additional file 2 Format: PDF

Title and description: The expected number of normal breast epithelial duct, ductal carcinoma in situ (DCIS), primary breast carcinoma and/or lymph node (LN) metastasis tissues (n).

Additional file 3 Format: PDF

Title and description: The clinico-pathological and survival data for patients.

Additional file 4 Format: PDF
Title and description: The oestrogen receptor (ER), progesterone receptor (PR) and human epidermal growth factor receptor-2 (HER2) status data for patients.

Additional file 5 Format: PDF
Title and description: Working with tissue microarray (TMA). Describes the steps involved in TMA construction, template editing, design computation, punch area selection and heat cycling.

Additional file 6 Format: PDF
Title and description: Product and supplier details for the antibodies used in this study.

Additional file 7 Format: PDF
Title and description: Immunohistochemistry (IHC) conditions for antigen detection.

Additional file 8 Format: PDF
Title and description: Detailed information on steps involved in using the program Distiller.

Additional file 9 Format: PDF
Title and description: Semiquantitative evaluation of immunohistochemical immunoreactivity and consolidation of TMA scoring.

FIG 1

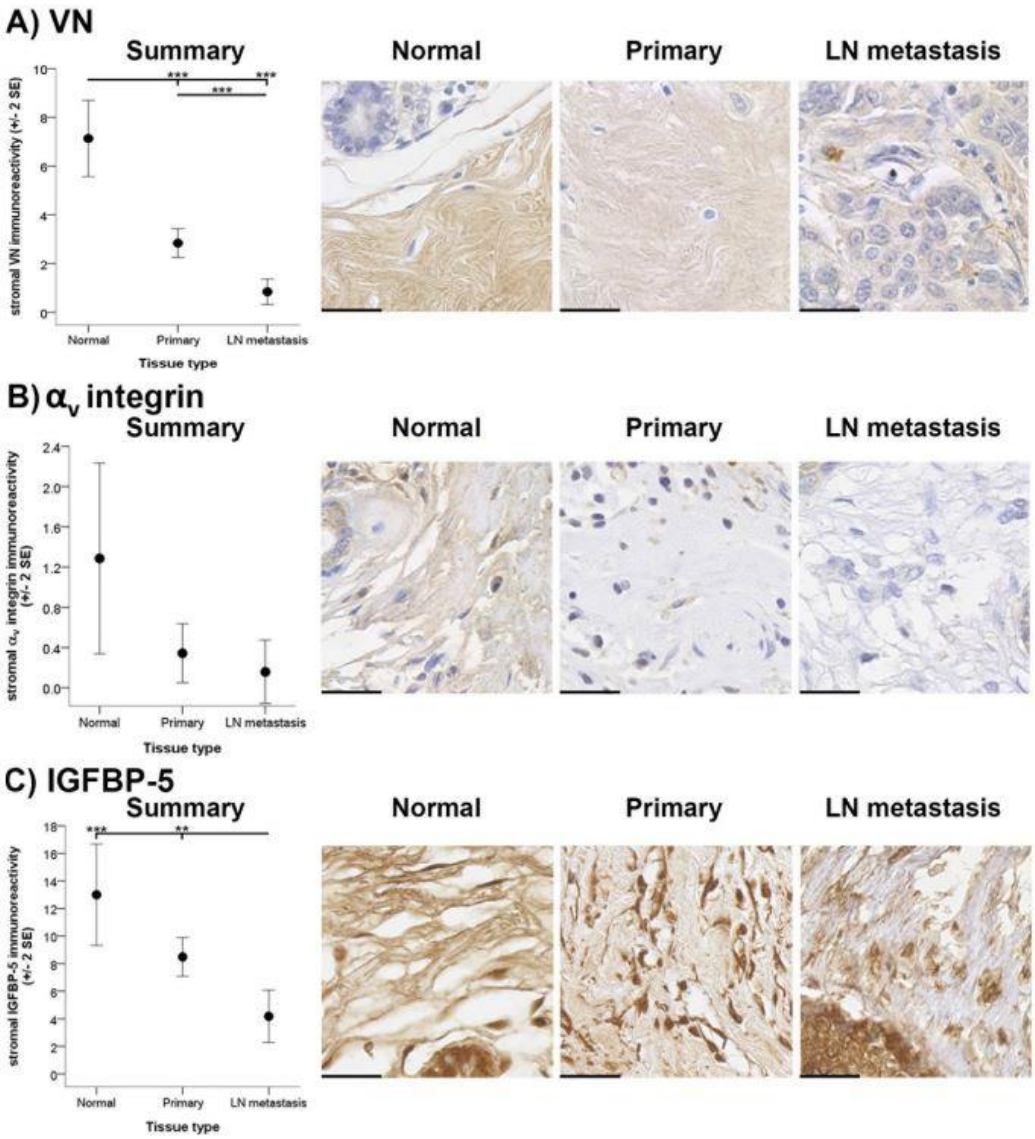

A) VN
Summary
Normal
Primary
LN metastasis
stromal VN immunoreactivity (+/- 2 SE)
Tissue type
B) αv integrin
Summary
Normal
Primary
LN metastasis
stromal αv integrin immunoreactivity (+/- 2 SE)
Tissue type
C) IGFBP-5
Summary
Normal
Primary
LN metastasis
stromal IGFBP-5 immunoreactivity (+/- 2 SE)
Tissue type

FIG 2

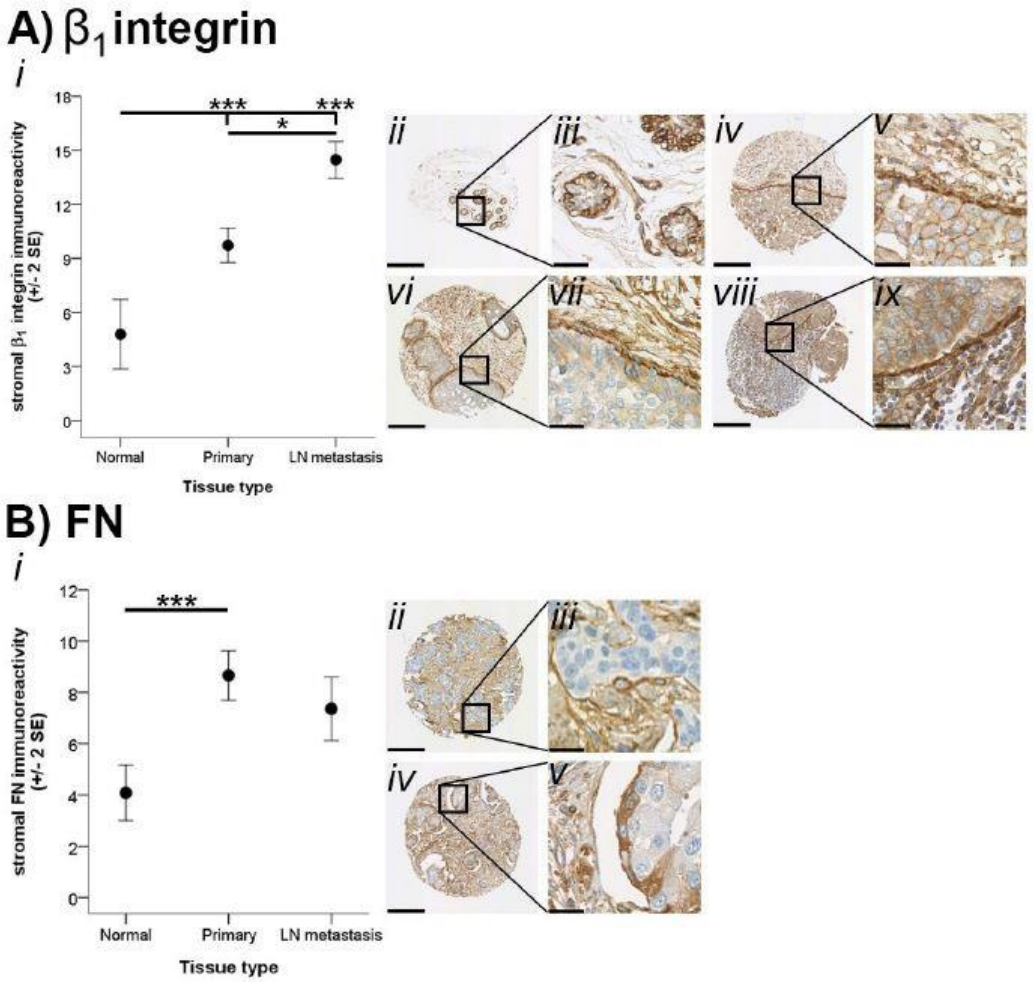

A) β1 integrin
i
stromal β1 integrin immunoreactivity (+/- 2 SE)
***
***
*
Normal
Primary
LN metastasis
Tissue type
ii
iii
iv
v
vi
vii
viii
ix
B) FN
i
stromal FN immunoreactivity (+/- 2 SE)
***
Normal
Primary
LN metastasis
Tissue type
ii
iii
iv
v

FIG 3

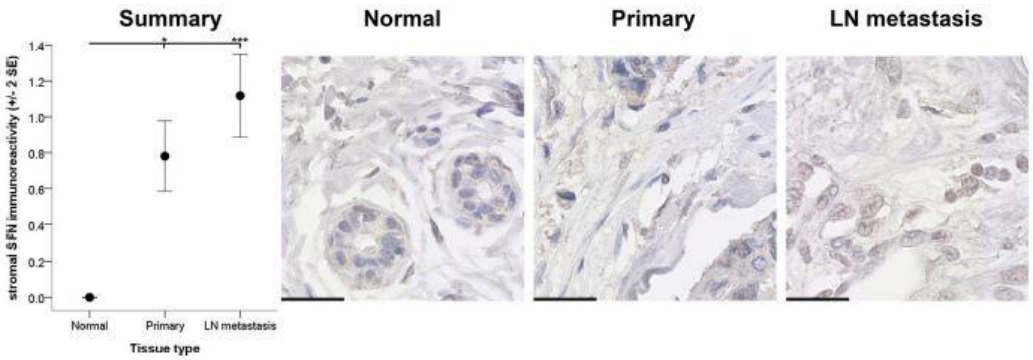

Summary
Normal
Primary
LN metastasis
stromal S FN immunoreactivity (+/- 2 SE)
Normal
Primary
LN metastasis
Tissue type

FIG 4

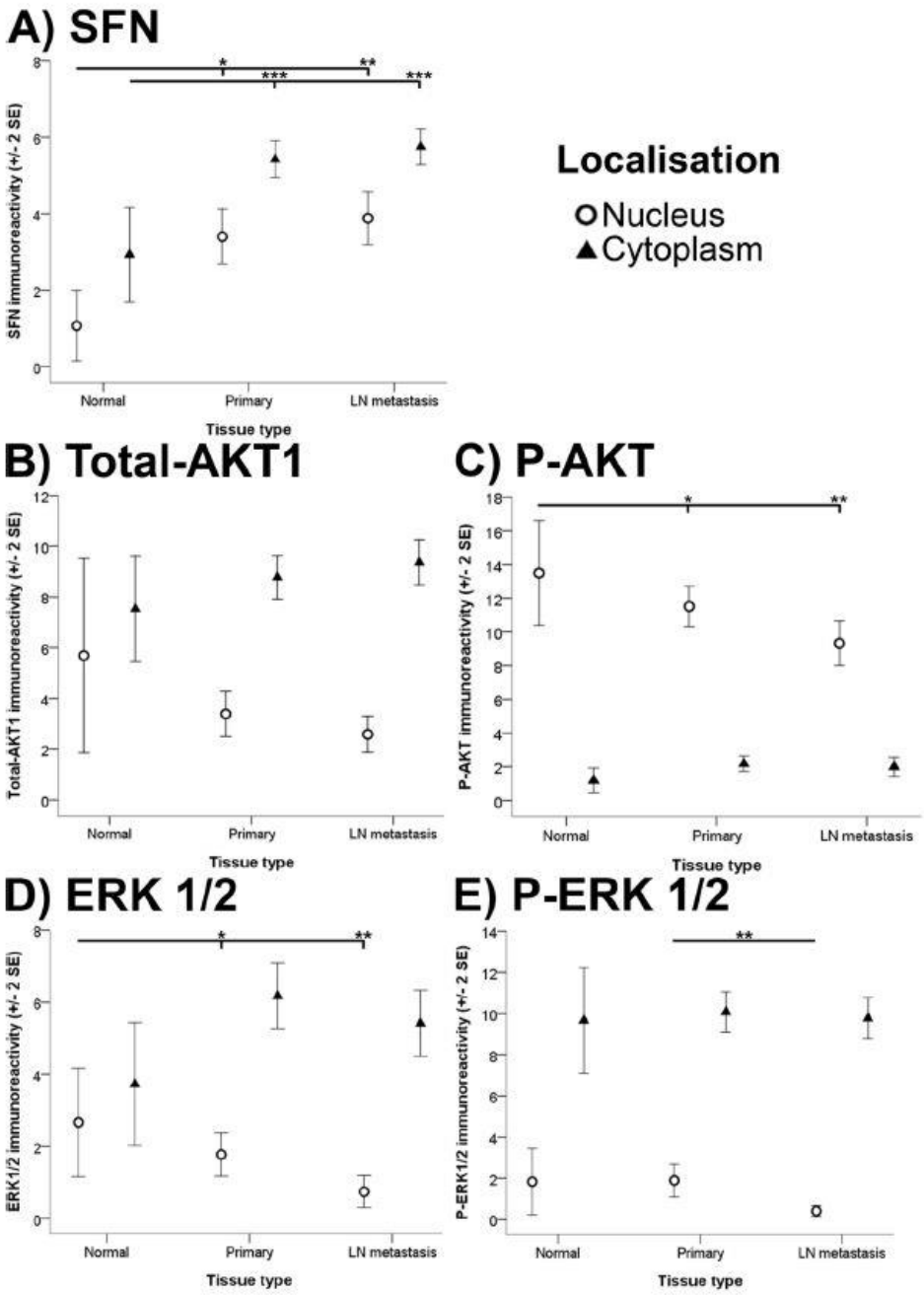

A) SFN
SFN immunoreactivity (+/- 2 SE)
Normal
Primary
LN metastasis
Tissue type
Localisation
Nucleus
Cytoplasm
B) Total-AKT1
Total-AKT1 immunoreactivity (+/- 2 SE)
C) P-AKT
P-AKT immunoreactivity (+/- 2 SE)
D) ERK 1/2
ERK1/2 immunoreactivity (+/- 2 SE)
E) P-ERK 1/2
P-ERK1/2 immunoreactivity (+/- 2 SE)

FIG 5

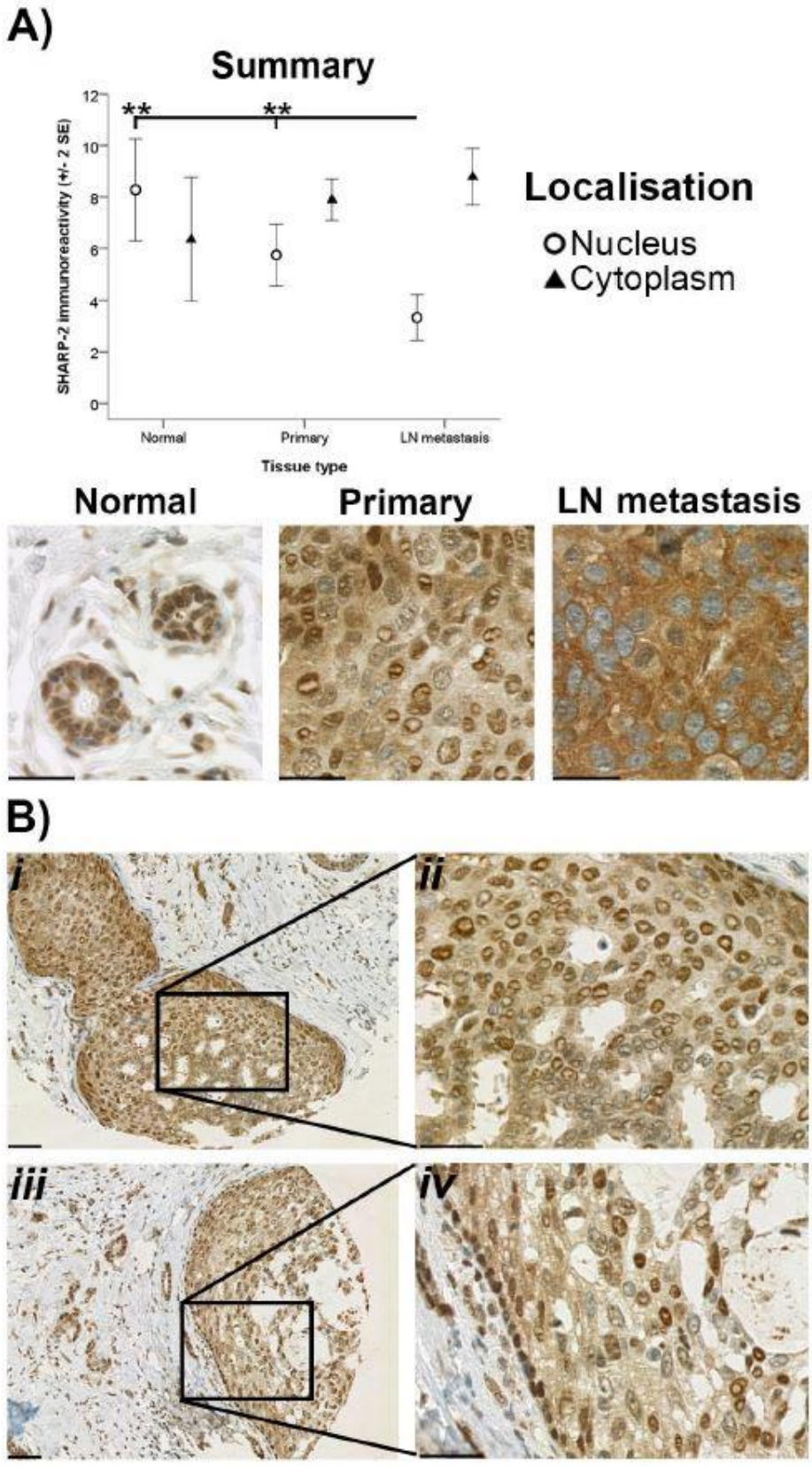
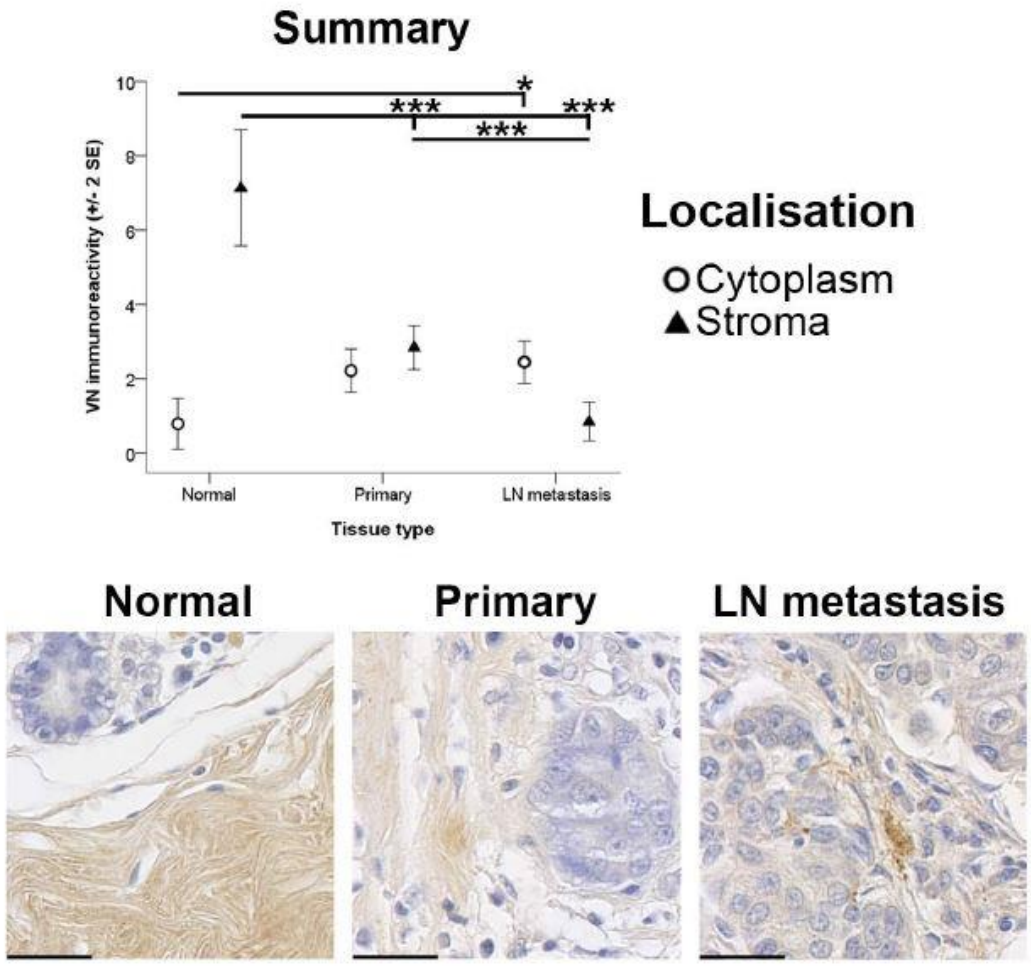

A)
Summary
**
**
SHARP-2 immunoreactivity (+/- 2 SE)
12
10
8
6
4
2
0
Normal
Primary
LN metastasis
Tissue type
Localisation
Nucleus
Cytoplasm
Normal
Primary
LN metastasis
B)
i
ii
iii
iv

FIG 6


Summary
VN immunoreactivity (+/- 2 SE)
10
8
6
4
2
0
*
***
***
***
Normal
Primary
LN metastasis
Tissue type
Localisation
Cytoplasm
Stroma
Normal
Primary
LN metastasis

Supplementary Table 1A. The actual number of normal breast epithelial duct, ductal carcinoma in situ (DCIS), primary breast carcinoma and/or lymph node (LN) metastasis tissues (n). N/A: not applicable.

TMA Name	Normal breast TMA cores (n)	DCIS TMA cores (n)	Primary cancer TMA cores (n)	LN metastasis TMA cores (n)
α_v integrin	7	3	32	19
β_1 integrin	14	5	84	68
CLDN1	12	7	82	66
ER	16	2	73	64
ERK1/2	15	2	84	67
FN	12	5	85	67
HER2	14	1	85	71
IGF-IR	14	1	34	21
IGF-IIR	16	6	79	69
IGFBP-5	6	N/A	35	23
P-AKT	16	2	86	68
P-ERK1/2	12	5	76	64
PR	11	4	71	65
SFN	14	6	82	68
SHARP-2	14	6	86	70
Total-AKT1	13	4	85	68
VN	23	2	86	68

Supplementary Table 1B. The expected number of normal breast epithelial duct, ductal carcinoma in situ (DCIS), primary breast carcinoma and/or lymph node (LN) metastasis tissues (n).

TMA name	Normal breast TMA cores (n)	DCIS TMA cores (n)	Primary cancer TMA cores (n)	LN metastasis TMA cores (n)
α_v integrin	12	0	38	25
β_1 integrin	24	10	88	76
CLDN1	24	10	88	76
ER	24	10	88	76
ERK1/2	24	10	88	76
FN	24	10	88	76
HER2	24	10	88	76
IGF-IR	12	0	38	25
IGF-IIR	24	10	88	76
IGFBP-5	12	0	38	25
P-AKT	24	10	88	76
P-ERK1/2	24	10	88	76
PR	24	10	88	76
SFN	24	10	88	76
SHARP-2	24	10	88	76
Total-AKT1	24	10	88	76

VN	24	10	88	76

Supplementary Table 2A. The clinico-pathological and survival data for patients. Total number (N) and proportion (%) of patients are indicated.

	Patients (N)	Proportion (%)
Age (years)		
< 40	4	4
40 - <50	27	30
50 - <60	25	27
60 - <70	22	24
70 - <80	11	12
80 - <90	1	1
Missing	1	1
Mean (± standard deviation)	55.91 (± 10.92)	
Tumour type (No/Yes/Missing)		
Ductal	6/79/6	7/87/7
Lobular	74/11/6	81/12/7
Tubular	83/2/6	91/2/7
Pleomorphic	83/2/6	91/2/7
Trabecular	84/1/6	92/1/7
Atypical medullary	58/1/32	64/1/35
Overall tumour grade		
1	3	3
2	15	16
3	59	65
Missing	14	15
Tumour size (mm) (maximum)		
≤ 10	4	4
11-20	21	23
21-30	26	29
31-40	15	16
41-50	7	8
51-60	3	3
61-70	3	3
71-80	1	1
81-90	1	1
91-100	1	1
Missing	25	27
Positive lymph nodes		
1-3	42	46
4-9	20	22
≥10	27	30
Missing	2	2
Lymph node ratio		
≤ 0.25	46	51
>0.25, ≤ 0.75	32	35
≥ 0.75	10	11
Missing	3	3
Survival status		
Deceased (due to disease)	44	48
Censored	0	0
Deceased (due to other causes)	1	1

	Patients (N)	Proportion (%)
Lost at date of death follow-up	36	40
Lost prior to follow-up	10	11

Supplementary Table 2B. The oestrogen receptor (ER), progesterone receptor (PR) and human epidermal growth factor receptor-2 (HER2) status data for patients. Total number (N) and proportion (%) of patients are indicated.

	Patients (N)	Proportion (%)
ER percentage		
0	21	23
1 - 10	3	3
11 -20	3	3
21 - 30	1	1
31 - 40	0	0
41 - 50	1	1
51 - 60	3	3
61 - 70	1	1
71 - 80	6	7
81 - 90	11	12
91 - 100	23	25
Missing	18	20
ER intensity		
0	21	23
1	12	13
2	19	21
3	21	23
Missing	18	20
PR percentage		
0	29	32
1 - 10	10	11
11 -20	4	4
21 - 30	4	4
31 - 40	1	1
41 - 50	6	7
51 - 60	2	2
61 - 70	5	5
71 - 80	3	3
81 - 90	2	2
91 - 100	5	5
Missing	20	22
PR intensity		
0	29	32
1	14	15
2	18	20
3	10	11
Missing	20	22
HER2 percentage		
0	46	51
1 - 10	7	8
11 -20	4	4
21 - 30	4	4
31 - 40	1	1
41 - 50	1	1
51 - 60	2	2
61 - 70	2	2
71 - 80	4	4

	Patients (N)	Proportion (%)
81 - 90	6	7
91 - 100	7	8
Missing	7	8
HER2 intensity		
0	46	51
1	16	18
2	7	8
3	15	16
Missing	7	7

Supplementary Materials and Methods 1.

Tissue microarray (TMA) construction: The Galileo TMA (CK3000) Tissue Microarrayer (Integrated Systems Engineering S.r.l., Milan, Italy) using the IseTMA software program (©BioRep, Milan, Italy) and the Beecher (ATA-27) Automated Tissue Arrayer (Beecher Instruments, Inc., Sun Prairie, Wisconsin, USA) were used to construct the TMAs. The methodologies listed below are those performed whilst using the Galileo TMA Tissue Microarrayer with the IseTMA software program. TMAs constructed by the MaCHR Laboratory using the Beecher (ATA-27) Automated Tissue Arrayer were utilised for this study following the manufacturer's instructions.

Tissue microarray (TMA) template editing: TMA templates were created following the TMA grid design phase. The dimensions of the blank, recipient FFPE blocks were determined and recorded in the IseTMA software program. The dimensions of the resulting TMAs were dependent on the physical dimensions of the recipient FFPE blocks. The geometry of the array (which included the required number of rows and columns) was defined using the following criteria: the block dimensions; X and Y margins; the needle diameter (0.6 mm); the number of spots; and the spacing (or the X and Y pitch) between the spots. For this study, 0.6 mm diameter cores were selected. Rows and columns of spacing gaps were included to further aid in the orientation of the TMA. Finally, the ES 111 Optiscan™ Upright Microscope Stage (Prior Scientific Inc., Rockland, Massachusetts, USA) was calibrated to ensure the precise positioning of the TMA within the centre of the recipient FFPE blocks. The template file created was saved as a .csv file.

Tissue microarray (TMA) design computation: Within the IseTMA software program, the TMA template file was opened and an appropriate TMA construction sequence selected. The FFPE blocks included in the study (commonly known as donor FFPE blocks) were

associated with one spot or multiple spots within the TMA. To achieve this, the patient AP codes were ‘linked’ to the individual TMA cores within the TMA template. Colour codes were used to identify TMA cores associated with normal breast epithelial ducts, DCIS, primary breast carcinoma and LN metastasis. The number of replicas of the created TMA was defined. To create duplicate TMAs, two recipient FFPE block copies were selected and codes created for each copy. Each of the operations listed above were performed remotely from the arrayer. The design file created was saved as a .csv file.

Tissue microarray (TMA) construction: Initially, the arrayer was calibrated and the TMA template and design files were opened. The TMA needle was calibrated to ensure the precise positioning of the paraffin blocks under the needles in the operation conditions. A high resolution ProgRes® C10plus camera (JENOPTIK Optical Systems GmbH, Jena, Germany) was utilised for TMA needle calibration. The two recipient FFPE blocks were inserted into the Upright stage. The IseTMA software program provided an operating flowchart to assist in TMA construction.

Tissue microarray (TMA) core punch area selection - Donor block and haematoxylin and eosin (H&E) slide alignment: Where possible, the cores were obtained from the leading edge of the tumour, thought to be where interactions between ligands in the ECM and the cancer cells were more likely to occur (Giovannucci 1999; Gladson et al. 1991; Tomasini-Johansson et al. 1994). This was achieved by inserting the donor FFPE blocks into the upright stage and choosing the donor TMA spot positions. To localise the sample, an H&E-stained tissue section was placed on top of its corresponding FFPE block. Following this, the pre-marked H&E slide was manually aligned with the donor block with reference to the monitor. This allowed for the accurate identification of the tumour edge. Once aligned, a number of donor spot positions, along the tumour’s edge (if applicable), were selected using a joystick. All of the punch positions for the donor FFPE block were saved at the same time. Subsequently, holes were prepared in the recipient FFPE blocks and sample cores were transferred from the donor FFPE block and inserted into the corresponding holes in the recipient FFPE blocks. This process was performed manually and repeated for each of the donor FFPE blocks.

Tissue microarray (TMA) heat cycling: To stabilise and prevent the loss of TMA cores during sectioning and immunohistochemistry (IHC) procedures a thin layer of wax was applied to the surface of all of the created TMAs containing the sample cores from the donor FFPE blocks. This was followed by a heat cycling step which involved heating the TMAs in a conventional laboratory oven at 30 degrees Celsius (°C) for short increments of time over 24 hours.

Supplementary Table 3. Product and supplier details for the antibodies used in this study. IgG: immunoglobulin G; mAb: monoclonal antibody; pAb: polyclonal antibody, and; N/A: not applicable.

Antibody target	Supplier	Supplier details	Product Number	Isotype	Source	Clone
IGF-IRβ	Santa Cruz Biotechnology®	Santa Cruz, California, USA	sc-713	IgG	Rabbit pAb	C-20
IGF-IIR	Santa Cruz Biotechnology®	Santa Cruz, California, USA	sc-25462	IgG	Rabbit pAb	H-300
IGFBP-5	R&D Systems®	Gymea, NSW, Australia	MAB875	IgG_{2b}	Mouse mAb	164503
VN	Epitomics®	Burlingame, California, USA	1730-1	IgG	Rabbit mAb	EP781Y
FN	Novocastra	North Ryde, NSW, Australia	NCL-FIB	IgG_1	Mouse mAb	568
α_v integrin	Calbiochem®	Gibbstown, New Jersey, USA	407286	IgG_1	Mouse mAb	272-17E6
β_1 integrin	Abcam®	Cambridge, Massachusetts, USA	ab3167	IgG_1	Mouse mAb	4B7R
Total-AKT1	Cell Signalling Technology®	Danvers, Massachusetts, USA	2967	IgG_1	Mouse mAb	2H10
P-AKT (Ser473)	Cell Signalling Technology®	Danvers, Massachusetts, USA	4051L	IgG_{2b}	Mouse mAb	587F11
ERK1/2	Cell Signalling Technology®	Danvers, Massachusetts, USA	4695	IgG	Rabbit pAb	137F5
P-ERK1/2 (Thr202/Thr204)	Cell Signalling Technology®	Danvers, Massachusetts, USA	9106	IgG_1	Mouse mAb	E10
CLDN-1	Invitrogen™	Mulgrave, VIC, Australia	18-7362	IgG	Rabbit pAb	N/A
SFN	Abcam®	Cambridge, Massachusetts, USA	110-02810	IgG_1	Mouse mAb	3C3
SHARP-2	Sigma-Aldrich®	Castle Hill, NSW, Australia	S8443	IgG	Rabbit pAb	N/A
ER	Ventana Medical Systems®	Tucson, Arizona, USA	790-4325	IgG	Rabbit mAb	SP1
PR	Ventana Medical Systems®	Tucson, Arizona, USA	790-4296	IgG	Rabbit mAb	1E2
HER2	Ventana Medical Systems®	Tucson, Arizona, USA	790-2991	IgG	Rabbit mAb	4B5

Supplementary Table 4. Immunohistochemistry (IHC) conditions for antigen detection. Incubation times are provided in minutes (mins). °C: degrees Celsius.

Antigen	Clone	Blocking non-specific background		Antigen Retreival		Antigen detection				Antigen probe	HRP-Polymer	Peroxidase Chromogen conditions
		Endogenous peroxidase	Protein	Solution	Conditions	System	Incubation time	Dilution	Diluent			
α_v integrin	272-17E6	5 min	10 min	Diva	40 min at 99°C with 30 min cooling	MACH 4™	30 min	1:50	Da Vinci	10 min	10 min	1 min
β_1 integrin	4B7R	Performed onboard		Protease 1	10 min	ultraView™	32 min	1:2500	Da Vinci	Performed onboard		
CLDN-1	N/A	Performed onboard		Borg	4 min at 125°C with 30 min cooling	ultraView™	32 min	1:100	Da Vinci	Performed onboard		
ER	SP1	Performed onboard				iView™	30 min	Pre-diluted antibody		Performed onboard		
ERK1/2	137F5	5 min	10 min	Not Performed		MACH 3™	60 min	1:150	Da Vinci	15 min	15 min	5 min
FN	568	Performed onboard		Reveal	4 min at 125°C with 30 min cooling	ultraView™	32 min	1:200	Da Vinci	Performed onboard		
HER2	4B5	Performed onboard				iView™	32 min	Pre-diluted antibody		Performed onboard		
IGF-IRβ	C-20	5 min	10 min	Diva	4 min at 125°C with 30 min cooling	MACH 4™	60 min	1:50	Da Vinci	15 min	15 min	1 min
IGF-IIR	H-300	Performed onboard		Borg	4 min at 125°C with 30 min cooling	ultraView™	32 min	1:50	Da Vinci	Performed onboard		
IGFBP-5	164503	5 min	10 min	Not Performed		MACH 4™	30 min	1:75	Da Vinci	10 min	10 min	1 min
P-AKT	587F11	Performed onboard		Reveal	4 min at 125°C with 30 min cooling	ultraView™	32 min	1:50	Da Vinci	Performed onboard		
P-ERK1/2	E10	5 min	10 min	Not Performed		MACH 4™	60 min	1:100	Da Vinci	10 min	15 min	2 min
PR	1E2	Performed onboard				iView™	30 min	Pre-diluted antibody		Performed onboard		
SFN	3C3	Not Performed	Not Performed	Reveal	4 min at 125°C with 30 min cooling	MACH 4™	60 min	1:50	Da Vinci	10 min	15 min	5 min
SHARP-2	N/A	Performed onboard		Reveal	4 min at 125°C with 30 min cooling	ultraView™	32 min	1:1600	Da Vinci	Performed onboard		
Total-AKT1	2H10	Performed onboard		Reveal	4 min at 125°C with 30 min cooling	ultraView™	32 min	1:50	Da Vinci	Performed onboard		
VN	EP781Y	5 min	10 min	Not Performed		MACH 3™	30 min	1:50	Da Vinci	10 min	10 min	4 min

Supplementary Materials and Methods 2.

Distiller, a secure, web based, flexible information management system, was used to facilitate the integration of clinical records, research data, digital TMA slides and different data types into a easy-to-use hierarchical database. Authorised consent was obtained from G.D.F. to gain access to the Brisbane Database of all PAH breast cancer patients.

Data uploading and browsing

The ‘Browse’ tab was used to upload ‘general’ patient data files (including patient, biopsy, survival and treatment data) into Distiller. The ‘Browse’ tab was also used to find all of the demographic, biopsy, treatment, follow-up, TMA and full face biomarker information, as well as the images and scoring information for all TMA cores for each patient.

Uploading general patient data files: Click on the ‘Browse’ tab > Click on the ‘Files’ tab > Upload required files.

Browsing patient information: Click on the ‘Browse’ tab > Click on the ‘Data’ tab > Upload required files.

Management of high throughput design and analysis of TMA data

The ‘OpTMA’ and ‘Administration’ tabs were used to manage high throughput design and analysis of TMA data within the Distiller framework. This included uploading, creating, editing, approving and/or managing the TMA spot groups, TMA scoring forms, TMA maps and TMA slides.

Creating TMA spot group: Click on the “Administration” Tab > Click on the “DB Admin Tab” Tab > Click on the “Create new data group” icon > Type “TMA” into the “Group type name” section > Select “TMA” in the “Parent” tab > Select “TMA Spots” in the “Table type” tab > Click on the “Submit” icon.

Editing TMA spot group attributes: The mandatory TMA spot group default attributes include: Spot ID; Treatment name; Slide ID; Row Number; Column Number and Block reference. Click on the “Administration” Tab > Click on the “DB Admin Tab” Tab > Click on the TMA spot group of interest > Select “Choice” in the “Type” tab > Add treatment/stains/biomarker types.

Creating TMA scoring form group: Click on the “Administration” Tab > Click on the “DB Admin Tab” Tab > Click on the “Create new data group” icon > Type “TMA Score Antibody

name" into the "Group type name" section > Select "TMA" in the "Parent" tab > Select "TMA scoring form" in the "Table type" tab > Click on the "Submit" icon.
Editing TMA scoring form attributes: Editing the name of a TMA scoring group: Click on the "Administration" Tab > Click on the "DB Admin Tab" Tab > Scroll down to the TMA scoring group of interest > Click on the 2nd column icon "Edit name" > Click on the 2nd icon "Save" button to save changes. Modifying and adding questions for a TMA scoring group: Click on the "Administration" Tab > Click on the "DB Admin Tab" Tab > Scroll down to the TMA scoring group of interest > Click on the 1st column icon "view attributes/questions' > Click on the "create new question" icon > Type in the question name > Select if question is unique > Select whether the question is mandatory > Select an appropriate data type > Click on the "Submit" icon. Creating choice attributes: Click on the "Administration" Tab > Click on the "DB Admin Tab" Tab > Scroll down to the TMA scoring group of interest > Click on the 1st column icon "view attributes/questions' > Click on the "Choice" button for the attribute of interest > Select the type of "Choice frequency" > Click on "Create new choice type" > Type in the Choice name > Type in the Choice value > Click on the "Edit/Save" icon > Choose the position of the choices.
Managing TMA maps: Uploading TMA maps: Click on the "OpTMA" Tab > Click on the "TMA Maps" Tab > Click on the "Upload" Tab to upload a TMA map from a Microsoft Excel™ spreadsheet. Editing and approving TMA maps: Click on the "OpTMA" Tab > Click on the "TMA Maps" Tab > Click on the TMA map of interest > Modify/Approve where necessary.
Managing TMA slides: Uploading TMA slides for dearraying: Click on the "OpTMA" Tab > Click on the "TMA slides" Tab > Click on tabs for the folders of interest which are located on the QUT server > Click on the link for the slide of interest > Select the TMA of interest in the "block reference" section > Click "TMA" for the "Group type name" section > Select the antibody of interest in the "treatment" section (add a new treatment name) > Select "x40" in the "Magnification" section > Select the relevant orientation (N.B. There were 8 possible scanning orientations for the slides) > Click on the "Submit" icon for dearraying to occur. Approving TMA slides: Click on the "OpTMA" Tab > Click on the "TMA slides" Tab > Expand the "Block Reference" tab to view the slides associated with this block reference > Click on the link under "Dearray comments" to view the predicted grid > Edit/Fix the predicted grid structures. Editing the predicted grid structures: The annotations (red circles) mark each core. Moving the cursor over each of the annotations will display the

row and column number of each core. To reposition, redraw, delete the annotation, and/or draw new cores, right-click on the annotation. The row & column numbers are assigned manually or automatically, by regridding the slide. Regridding attempts to reconstruct the TMA grid using all the annotations including any newly created cores. To save the edited grid structure press “F5”. Recheck the dearray comments and modify if necessary. If the predicted grid matches the map, the slide can be approved. Once the grid structure has been approved, an appropriate scoring form and list of scorers are selected.
Scoring the IHC stained TMAs: Scoring TMAs: Click on the “OpTMA” Tab > Click on the “score TMAs” Tab > Click on the appropriate block reference tab > Click on the “Score” tab to score the TMA slide > Fill in the blank spaces or highlight the appropriate choice in the TMA scoring sheet that appears of the screen > Click “Next” to proceed to the next TMA core (if necessary, click on “Previous” to go back to the previous TMA core) > Click the “F5” button to save the scoring data. Reviewing and Re-scoring TMAs: Click on the “OpTMA” Tab > Click on the “score TMAs” Tab > Click on the appropriate block reference tab > Click on the “Review” tab to review the scored TMA slide or to re-score the TMA slide > Review or re-score the TMA cores by filling in the spaces provided or highlighting the appropriate choice in the TMA scoring sheet that appears of the screen > Click “Next” to proceed to the next TMA core (if necessary, click on “Previous” to go back to the previous TMA core) > Click the “F5” button to save changes.

Creating searches to extract patient, biopsy, survival, treatment and TMA scoring data
The ‘Search’ tab was used to create and re-run old searches to extract the patient, biopsy, survival, treatment and TMA scoring data of interest from Distiller. Only the attributes or data of interest were selected to be displayed in the search results to limit the extent of data extracted for analysis.
Creating searches: Click on the “Search” Tab > Click on the “Constraints” Tab > Scroll down to the Data groups section > Click on the “Patient Information” Tab > Type “P” into the “Patient ID” section > Click on the “TMA score ….” Tab for the protein of interest > Tick the appropriate protein in “Biomarker” section > Click on the “TMA” Tab > Scroll down to the “Block Reference” section > Type in the appropriate TMA design name of interest > Click on the “Output” tab > Scroll down to the Data groups section > Click on the “Patient Information” tab > Select “Select all” to highlight all fields in this data group > Click on the “Biopsy Information” tab > Select “Select all” to highlight all fields in this data group > Click

on the “Pathology” tab > Select “Select all” to highlight all fields in this data group > Click on the “Treatment Information” tab > Select “Select all” to highlight all fields in this data group > Click on the “Follow up Information” tab > Select “Select all” to highlight all fields in this data group > Click on the “TMA” tab > Select “Select all” to highlight all fields in this data group > Click on the “TMA score ….” tab for the protein of interest > Select “Select all” to highlight all fields in this data group > Click on the “Results” Tab to view the search results > Click “Download” to download the search results in a CSV format > Click “Save Search Parameters” to save the search created.

Supplementary Materials and Methods 3.
Semiquantitative evaluation of immunohistochemical immunoreactivity and consolidation of TMA scoring

A biomarker scoring form was displayed alongside each TMA core for completion. A standard biomarker scoring form was used for all protein targets investigated. Modified biomarker scoring forms were used for ER, PR and HER2 (scoring forms not shown). Specific questions were compulsory. Completion of the scoring forms for each TMA core was required prior to moving on to the next TMA core. Briefly, the TMA scorers were asked to record: their initials; the type of tumour present (provided as a choice between normal breast tissue, DCIS, primary tumour, metastatic tumour tissue from the LN and missing tissue); the amount of tumour present (as a percentage); the quality of the IHC staining (provided as a choice between yes or no); the percentage of cells with cytoplasmic staining (as a percentage); the intensity of the cytoplasmic staining (provided as a choice between 0, 1, 2 or 3); the percentage of cells with nuclear staining (as a percentage); the intensity of the nuclear staining (provided as a choice between 0, 1, 2 or 3); the percentage of cells with membrane staining (as a percentage); the intensity of the membrane staining (provided as a choice between 0, 1, 2 or 3); the percentage of stromal staining (as a percentage); and the intensity of the stromal staining (provided as a choice between 0, 1, 2 or 3).

IHC TMA scoring data was evaluated semiquantitatively using five scoring methods. Details on these scoring methods are listed below.

Presence: Data were expressed as the proportion of the total number of samples in which protein immunoreactivity was present within the normal breast tissue, DCIS, primary tumour or LN metastatic tumour tissue.

Intensity: Data were expressed as a score between 0 and 3, corresponding to the presence of negative, weak, intermediate, or strong staining, respectively, within the normal breast tissue, DCIS, primary tumour or LN metastatic tumour tissue.

Percentage: Data were expressed as the proportion of cells or stroma staining positively within the normal breast tissue, DCIS, primary tumour or LN metastatic tumour tissue.

Percentage class: Data were expressed as scores between 1 and 6 (1 = 0-4%; 2 = 5-19%; 3 = 20-39%; 4 = 40-59%; 5 = 60-79%, and; 6 = 80-100%), corresponding to the proportion of cells staining positively within the normal breast tissue, DCIS, primary tumour or LN metastatic tumour tissue.

Quickscore (Q) score: Data were expressed as scores between 0 and 18 using a multiplicative Q system (Detre et al., 1995b); this involves determining the staining intensity in addition to the proportion of cellular staining. The multiplicative method allows for a “no staining” category.

The exported .csv files from Distiller were converted to Microsoft Excel™ spreadsheets for each search performed. Each of the spreadsheet rows corresponded to a TMA core (with a specific tissue type). The variables recorded in the scoring form and the clinico-pathological data was represented as separate columns in the spreadsheet. Duplicate rows represented TMA cores originating from the same tissue sample and placed in duplicate TMAs. If there were multiple TMA scorers then duplication rows would be present for each TMA scorer. If a specific tissue sample was placed into more than one set of duplicate TMAs than duplicate rows would be present for each set of duplicate TMAs. These duplicate rows were consolidated into one row (if possible); the method of consolidation for each variable is listed below. For statistical data analysis purposes, the consolidated rows (each with a specific tissue type) were split up into four rows, each representing the results for a specific cellular localisation.

Initials: If there was only one TMA scorer, the initials of that scorer was recorded in the consolidated row. If there was more than one person who scored the TMA slide for IHC immunoreactivity for a particular protein than the initials of all of the TMA scorers were recorded in the consolidated row.

The type of tumour present: If the type of tissue was the same across the duplicate rows then that type of tissue was recorded in the consolidated row. If there was more than one type of tissue then a consolidated row was required for each tissue type represented.

The amount of tumour present: A mean percentage was calculated based on the proportion of the tissue in the TMA core which contained primary or LN metastatic tumour tissue.

The quality of the immunohistochemical immunoreactivity: TMA cores with unsatisfactory staining were discarded, while TMA cores with satisfactory immunoreactivity were consolidated (by the methods mentioned in this section).

The percentage of cells with nuclear, cytoplasmic and membrane immunoreactivity: A mean percentage was calculated based on the proportion of cells, within the normal breast tissue, DCIS, primary tumour or LN metastatic tumour tissue which contained nuclear, cytoplasmic and/or membrane immunoreactivity.

The percentage of stromal immunoreactivity: A mean percentage was calculated based on the proportion of the stromal cells and ECM (which will be referred to as the ‘stroma’), surrounding the normal breast tissue, DCIS, primary tumour or LN metastatic tumour tissue which contained immunoreactivity.

The intensity of the nuclear, cytoplasmic, membrane and stromal immunoreactivity: The highest intensity score was recorded in the consolidated row.

Chapter 9

STIRZAKER, C., ZOTENKO, E., SONG, J. Z., QU, W., NAIR, S. S., LOCKE, W. J., STONE, A., ARMSTONG, N. J., ROBINSON, M. D., DOBROVIC, A., AVERY-KIEJDA, K. A., PETERS, K. M., FRENCH, J. D., STEIN, S., KORBIE, D. J., TRAU, M., FORBES, J. F., SCOTT, R. J., BROWN, M. A., FRANCIS, G. D. & CLARK, S. J. 2015. Methylome sequencing in triple-negative breast cancer reveals distinct methylation clusters with prognostic value. Nat Commun, 6, 5899.

9.1.1. Introduction

Epigenetic alterations were evaluated in a breast cancer cohort to determine if these biomarkers could be used to discriminate normal breast tissue from breast carcinoma; lymph node positive from lymph node negative patients and to identify different prognostic groups.

9.1.2. Results

Abstract

Epigenetic alterations in the cancer methylome are common in breast cancer and provide novel options for tumour stratification. Here we perform whole genome methylation capture sequencing on small amounts of DNA isolated from formalin- fixed, paraffin-embedded tissue from triple negative breast cancer (TNBC) and matched normal samples. We identify differentially methylated regions (DMRs) enriched in promoters associated with transcription factors binding sites and DNA hypersensitive sites. Importantly, we stratify TNBCs into three distinct methylation clusters associated with better or worse prognosis and identify 17 DMRs that show a strong association with overall survival, including DMRs located in the Wilms Tumour 1 (WT1) gene, bidirectional-promoter and anti-sense WT1-AS. Our data reveals coordinated hypermethylation can occur in ER-ve disease and that characterising the epigenetic framework provides a potential signature to stratify TNBCs. Together our findings demonstrate the feasibility of profiling the cancer methylome with limited archival tissue to identify regulatory regions associated with cancer.

Introduction

Triple negative breast cancers (TNBCs) comprise a heterogeneous group of cancers with varying prognoses, presenting a challenge for effective clinical management. TNBC is clinically defined by the absence of estrogen (ER) and progesterone receptor (PR) expression, and neither overexpression nor amplification of human epidermal growth factor receptor 2 (HER2) [1, 2]. TNBC represents approximately 15-20% of all newly diagnosed breast cancer cases and is generally associated with high risk of disease recurrence and shorter overall survival compared to non-TNBC[3]. Broadly, TNBC patients can be categorized into two distinct groups; those that succumb to their disease within 3-5 years regardless of treatment, and those that remain disease free to the extent that their

overall survival exceeds that of non-TNBC patients (i.e. approximately >8 to 10 years post-diagnosis)[4, 5]. Currently, methods by which TNBC patients are stratified into high- and low-risk subgroups remain limited to staging by clinicopathological factors such as tumour size, level of invasiveness and lymph node infiltration. However, unlike other breast cancer subtypes, TNBC outcome is less closely related to stage[6]. Thus, there is a clear need to identify a robust method by which TNBC patients can be stratified by prognosis, to enable more informed disease management.

Current efforts to stratify early breast cancer prognosis have primarily focused on multi-gene expression signatures and all have received varying degrees of acceptance (reviewed in [7]). In addition to multi-gene expression assays, DNA methylation signatures are being assessed as potential molecular biomarkers of cancer[8]. A number of studies have documented aberrant methylation events in breast carcinogenesis and identified specific DNA methylation biomarkers that have significant diagnostic and prognostic potential[9, 10, 11, 12]. Several studies have also identified DNA methylation signatures that can distinguish between breast cancer subtype[13, 14, 15, 16], and others that may be predictive of treatment response[17, 18, 19].

Despite growing interest in the prognostic significance of DNA methylation in breast cancer, there have been no studies specifically investigating the DNA methylation profile of human TNBC and its association with disease outcome. Here we carry out genome-wide DNA methylation profiling of formalin-fixed paraffin-embedded (FFPE) triple-negative clinical DNA samples, using affinity capture of methylated DNA with recombinant methyl-CpG binding domain of MBD2 protein, followed by next generation sequencing (MBDCap-Seq)[20, 21]. This high-resolution technique allows for genome wide methylation analysis of CpG rich DNA[22, 23]. Using MBDCap-Seq, we identify regional methylation profiles specific to TNBC, which we validate using methylation data extracted from TCGA breast cancer cohort[13]. Importantly, we also report the first potential prognostic methylation signature of survival, specific for TNBC that now warrants further study in larger cohorts.

Results

Genome Coverage of MBDCap-Seq

To delineate regions assayable with MBDCap-Seq we first profiled fully methylated (CpG methyltransferase SssI treated blood sample) DNA. Computational analysis of SssI MBDCap-Seq revealed that MBDCap-Seq can robustly assess the methylation status of 230,655 regions spanning a total of 116Mbp, comprising 5,012,633 CpG dinucleotides, or approximately 18% of the total number of CpG sites in the human genome (see Methods; Supplementary Fig. 1A). The assayed CpG sites span 91% of all CpG islands; 84% CpG island shores; 72% RefSeq promoters; 38% introns and 31% exons. We next compared coverage of MBDCap-Seq with the Illumina HumanMethylation450K (HM450K) array (Supplementary Fig. 1B) and found that MBDCap-Seq interrogates an additional 4,740,327 CpG sites as compared to the high-density HM450K array.

A major advantage of the MBDCap-Seq method is the ability to interrogate regional blocks of hypermethylation, that is methylation spanning consecutive CpG sites, which commonly occurs in cancer. We compared regional MBDCap-Seq coverage to coverage of HM450K arrays (Supplementary Fig. 1A) and found that while MBDCap-Seq and HM450K arrays have similar regional coverage of CpG islands (91% vs. 81%) and RefSeq promoters (71% vs. 83%), MBDCap-Seq regional coverage of shores (77% vs. 28%), enhancers (12% vs. 2%) and insulators (11% vs. 1%) is much greater, highlighting the potential advantage of MBDCap-Seq in screening novel functional regions of the cancer methylome.

To determine if MBDCap-Seq can also provide accurate methylation analysis from FFPET DNA, we compared DNA methylation profiles from DNA isolated from fresh frozen (FF) and FFPET of matching tumour and lymph node samples. We show that MBDCap-Seq from FFPET provides equivalent methylation to FF DNA (Pearson Correlation Coefficient of 0.95 and 0.86, respectively) (Supplementary Fig. 2A) and that MBDCap-Seq and HM450K array performed on the same FFPET tumour and lymph node DNA show high concordance (0.79 & 0.77 respectively) (Supplementary Fig. 2B-D). We also show that there are regions uniquely covered by MBDCap-Seq for example at enhancers and insulator elements (Supplementary Fig. 2E and F).

Identification and validation of differentially methylated regions in TNBCs

To identify differentially methylated regions (DMRs) in TNBCs we first profiled FFPET DNA using MBDCap-Seq from a discovery cohort of 19 Grade 3 TNBCs tumour and 6 matched normal samples (Supplementary Table 1) and analysed the data with a novel computational pipeline for comparative statistical analysis of MBDCap-Seq samples (see Methods; Supplementary Fig. 3). We identified 822 hypermethylated and 43 hypomethylated statistically significant DMRs (FDR <0.05), harboring 64,005 and 623 CpG sites respectively, compared to matched normal samples (see Fig. 1A & 1B; Supplementary Data 1) and validated sample specific differential methylation using Sequenom DNA methylation analysis (Supplementary Fig. 4). Next, we determined the genomic location of the DMRs and found that CpG islands, CpG island shores and promoters are significantly over-represented (hyper- geometric test; p-value < 0.0001) in the 822 hypermethylated regions and under- represented in the 43 regions of hypomethylation (Fig. 1C; See Methods). Notably, ChromHMM annotated HMEC promoters[24] and polycomb repressed regions were also significantly enriched (hyper-geometric test; p-value << 0.001) for gain of methylation in the breast cancer samples. Finally, we validated example DMRs in an independent cohort of 31 TNBCs and 15 normal breast samples and a panel of cell lines (Supplementary Table 2). We performed Sequenom methylation analysis on 5 of the 822 hypermethylated regions spanning the CpG island promoters of NPY, FERD3L, HMX2, SATB2 and C9orf125 (Supplementary Figure 5). The levels of methylation detected in the normal samples were uniformly low, whereas the 5 DMRs showed striking hypermethylation in the TNBCs (Fig. 1D), and 24 breast cancer cell lines (Fig. 1E).

Functional Characterization of Genes with Promoter Hypermethylation

To predict the potential functional significance of the 822 DMRs identified in the TNBC, we first determined which regions overlapped with promoters and genes and found that our DMRs were associated with 513 RefSeq promoters, which corresponded to 308 genes (Supplementary Data 2). We used the DAVID functional annotation tool[25] to annotate this set of genes. Visualization of statistically significantly (FDR < 0.05) over-represented gene sets revealed two largely non- overlapping groups of genes[26](See Methods; Supplementary Fig 6; Supplementary Table 3). One group is annotated with keywords “DNA-BINDING", "TRANSCRIPTION", "TRANSCRIPTION REGULATION", "HOMEOBOX", "DEVELOPMENTAL PROTEIN", and "DIFFERENTIATIONS" and

contains around 100 genes, mostly transcription factors, such as BARHL2, DLX6, OTX2, RUNX1T1 and TAC1. The second group is annotated with keywords "SIGNAL", "CELL MEMBRANE", "TRANSDUCER", "GLYCOPROTEIN", "G-PROTEIN COUPLED RECEPTOR" and contains genes involved in signaling pathways such as ADRB3, GHSR, NPY and ROBO3.

To determine if promoter hypermethylation was potentially involved in gene silencing we examined TCGA expression data for the 308 genes affected by promoter hypermethylation (see Methods for analysis of TCGA expression data for TNBC samples; 89 tumour and 8 matched normal samples). We found that genes with promoter hypermethylation are enriched in down-regulated genes (71 out of 245 genes for which expression data is available are down-regulated; hyper-geometric test; FC 1.73; p-value << 0.001) and are depleted in up-regulated genes (28 out of 245 genes are up-regulated; hyper-geometric test; FC 0.53; p-value << 0.001) (Fig. 1F).

To identify potential driver events we overlapped the 308 hypermethylated genes with genes recurrently mutated in breast cancer in TCGA[13] (Fig. 1G). We found that out of 308 genes with promoter methylation, 51 are mutated (hyper-geometric test; FC of 1.92; p-value << 0.001) and 12 (C9orf125, COL14A1, ENPP2, ERG2, PLD5, ROBO3, RUNX1T1, SEMA5A, TBX18, TSHZ3, ZBTB16, and ZNF208) are both mutated and down-regulated. Of these, both ROBO3 and SEMA5A are part of the axon guidance pathway recently implicated in tumour initiation and progression[27, 28]. Interestingly, promoter hypermethylation affects, in total, seven members of the axon guidance pathway (CRMP1, GDNF, GFRA1, MYL9, ROBO1, ROBO3 and SEMA5A) with four members (GFRA1, MYL9, ROBO3, and SEMA5A) being down-regulated.

Differentially Methylated Regions Specific to TNBCs

We next asked if any of the 822 DMRs were also found in ER-ve or ER+ve breast cancer. We used TCGA breast cancer methylation cohort, which comprises HM450K data for 354 ER+ve and 105 ER-ve breast tumours (73 of which are TNBCs) and 83 normal breast samples (see Methods for analysis of TCGA methylation data). Of the 822 DMRs regions identified in the MBDCap-seq methylation discovery set, 770 DMRs are interrogated by a total of 4,987 HM450K probes from the TCGA data set. We found that while the majority of

these probes are not methylated in breast normal tissue they were hypermethylated to various degrees in both ER+ve and ER-ve breast cancers (Fig. 2A). Both ER+ve and ER-ve subtypes also contained samples with minimal methylation across all probes, as well as those that displayed extensive methylation more representative of a CpG island methylator phenotype (CIMP)[29].

Next we asked if any of the DMRs were TNBC specific. Out of 4,987 HM450K probes, we found that 5% (282/4,987) were significantly hypermethylated in TNBCs (t-test; mean diff. methylation>10%; p-value<0.05) compared to the ER+ve tumours and the rest of the ER-ve tumours. Using methylation values of 282 TNBC-specific probes, we are able to classify tumour samples in the TCGA HM450K cohort into TNBCs and non-TNBCs with sensitivity of 0.72 sensitivity, specificity of 0.94 and AUC 0f 0.90 (Fig. 2B). From the 282 TNBC specific probes we identified 36 TNBC specific regions (harbouring at least 3 or more 450K TNBC-specific probes) that primarily overlap promoters and/or gene bodies (Supplementary Table 4; Supplementary Fig. 7). The regions predominantly overlap genes encoding zinc fingers and transcription factors and intergenic regions that are commonly marked by polycomb in HMECs. An example of two such TNBC specific regions, are located in the promoters of genes encoding zinc finger proteins ZNF154 and ZNF671 on chromosome19 (Fig. 2C). Both promoters have low methylation levels in normal breast and increased levels of methylation in TNBC samples as compared to ER+ve cancer. The distribution of expression values mirrors the methylation status, with normal samples showing the highest levels of expression and TNBC tumours showing the lowest levels of expression (Fig. 2D), suggesting silencing by methylation of both ZNF154 and ZNF671 in TNBC tumours.

DNA methylation profile can stratify TNBC

To identify DMRs that potentially stratify TNBCs we used unsupervised cluster analysis on methylation of the 4,987 HM450K probes and identified three distinct groups of TNBC tumours from the TCGA data sets (Fig 3A; see Methods). Survival analysis revealed that the largely hypomethylated cluster (blue cluster) was associated with better prognosis as compared to the other two more highly methylated clusters (orange and red clusters) (Fig. 3B). In particular, the medium methylated cluster (orange cluster) comprises samples with the worst prognosis (Cox proportional hazards model; HR 8.64; p-value

0.005) as compared to the good prognosis TNBC cluster (blue cluster). Moreover there was no association between the induced clusters and survival for ER+ve or non TNBC samples (Supplementary Fig 8).

Next, we determined to what extent regional methylation stratify TNBCs into good and bad prognosis groups. Survival analysis identified 190 probes that were associated with survival in TCGA TNBC samples (Cox proportional hazards model; p-value < 0.05) in both univariate and multivariate analyses (see Methods). We observed regional association (at least three concordantly located survival probes) for 17 regions; 14 genomic regions with poor survival and 3 genomic regions for good survival (Table 1; Supplementary Fig. 9). Each of the individual Kaplan Meier plots for individual CpG sites in each region showed excellent survival separation, highlighting the potential value as prognostic biomarkers (Fig. 3C-E; Supplementary Fig. 9). The genomic location of these regions vary with four regions located in a promoter (SLC6A3, C6orf174, WT1-AS and ZNF254), seven in the gene body only (DMRTA2, LHX8, WT1, WT1-AS, HOXB13, ECEL1, SOX2-OT) and five in intergenic regions (Table 1). Interestingly, with the exception of the region encoded by chr10: 102,409,068-102,409,766, all prognostic regions overlap DNase1 hypersensitive sites (ENCODE), are marked with a polycomb signature in HMEC cells and many contain numerous conserved transcription factor binding sites (TRANSFAC®,[30]) (Table. 1). Furthermore, we show that the average level of methylation of CpG sites in the 17 potential prognostic regions is higher in the two poor survival groups and is lower in the normal and low risk groups (Supplementary Fig. 10).

A striking example of regional hypermethylation across consecutive CpG probes that shows statistical significance as a prognostic marker of survival are the DMRs spanning the bidirectional promoter and gene bodies of WT1 gene and its antisense counter-part, WT1-AS (Fig. 3F). Wilms tumour protein (WT1) is a zinc finger transcription factor over-expressed in several tumour types including breast (reviewed in [31]). We observe an association between high level of methylation in chr11-11623 and chr11-1210, regions spanning the gene bodies of WT1 and WT1-AS respectively, and poor survival in TCGA TNBC cohort (Fig. 3F). Moreover, increased levels of methylation in these regions are also associated with increased expression of WT1 (chr11-11623) and WT1-AS (chr11-1210) in TNBC patients (Supplementary Fig. 11). Conversely, we observe that TNBC patients with

high methylation in chr11-4047, a region spanning bi-directional promoter of WT1 and WT1-AS, survive longer than TNBC patients with low methylation in this region.

Table 1: Summary of 17 DMRs associated with overall survival in TCGA TNBC.

Chr	Start	End	RefSeq Location	CpG Island	CpG Shore	Dnase Hypersensitive Site	Conserved Transcription Factor Binding Sites (Z score cutoff = 2.33)	ChromHMM HMEC Polycomb	Prognostic Probes	Gene Function/Location
Poor Prognosis										
chr1	50,658,646	50,659,783	DMRTA2*	YES	NO	YES	GRα	YES	3	Doublesex- And Mab-3-Related Transcription Factor
chr1	75,368,128	75,368,976	LHX8*	YES	YES	YES	n/a	YES	3	Homeobox protein Lhx8
chr10	102,409,068	102,409,766		YES	YES	NO	STAT1β, NK-κB, CREB, NF-Y, CEBPα	YES	5	Intergenic region, ChromHMM polycomb marked
chr11	32,404,535	32,407,465	WT1*	YES	YES	YES	EGR1, EGR2, EGR2, NF1, LMO2, RFX1, MIF1, CREB, cJUN, ATF, ATF2	YES	5	Wilms Tumor protein, transcription factor
chr11	32,416,010	32,417,947	WT1-AS*	YES	YES	YES	n/a	YES	4	Wilms Tumor protein, antisense transcipt
chr13	27,398,788	27,401,867		YES	YES	YES	n/a	YES	4	Intergenic region, ChromHMM polycomb marked
chr14	56,330,541	56,332,135		YES	YES	YES	USF1, MAX1, c-MYC	YES	3	Intergenic region, ChrommHMM polycomb
chr17	44,159,065	44,159,578	HOXB13*	YES	YES	YES	HSF1, HSF2	YES	3	Homeobox gene family, transcription factor
chr2	233,058,433	233,060,592	ECEL1*	YES	YES	YES	NRSF, PAX2, STAT5A, YY1, AHR, GATA2, AP2	YES	3	Zinc-containing typeII integral-membrane protein
chr3	182,923,564	182,924,686	SOX2-OT*	NO	NO	YES	n/a	YES	4	Non-protein coding RNA gene.
chr5	1,498,811	1,499,696	SLC6A3◆	YES	YES	YES	n/a	YES	3	Neurotransmitter reporter
chr6	27,620,848	27,621,582		NO	NO	YES	n/a	NO	3	ChromHMM promoter marked, Intergenic region
chr6	127,881,341	127,882,455	C6orf174*◆	NO	YES	YES	STAT5A, FOXC1	YES	6	Chromosome 6 Open Reading Frame SOGA3 protein coding region
chr7	121,726,837	121,728,266		YES	YES	YES	CHX10	YES	4	ChromHMM polycomb marked
Good Prognosis										
chr11	32,413,697	32,415,714	WT1/WT1-AS◆	YES	YES	YES	E47, AP4, c-MYC, ARNT	YES	5	Bidirectional Promoter of WT1/WT1-AS transcription factor
chr19	24,061,637	24,062,272	ZNF254*◆	NO	NO	YES	n/a	NO	4	Zinc finger protein,transcriptional regulation

chr22	44,641,414	44,642,542		YES	YES	YES	n/a	YES	3	Intergenic region, ChromHMM promoter
*Gene Body		◆Promoter								

Discussion

The prognostic stratification of TNBC patients remains one of the most significant challenges in breast cancer research. While current efforts have primarily focused on the development of multi-gene expression classifiers to inform patient outcome, here we demonstrate the significant prognostic potential of DNA methylation biomarkers for stratification of TNBCs. We performed genome-wide DNA methylation profiling on TNBC and identified novel regions of differential methylation, and validated regions specific for TNBC using TCGA methylation data as an independent cohort. Strikingly, unsupervised cluster analysis of DMRs stratified TNBC patients into populations of high, medium or low risk disease outcome. Additionally, using both univariate and multivariate Cox proportional hazard models, we identified 17 DMRs significantly associated with TNBC patient survival ($p<0.05$). Critically, our classifiers paralleled the biologically relevant time-dependent pattern of patient outcome, whereby TNBC patients are most vulnerable to disease associated death within the first 5 years following diagnosis; highlighting their potential use as a valuable prognostic application.

The DNA methylation aberrations we identified in the TNBC samples follow specific patterns common to many cancer types[32]. For instance, hypermethylation events were localized to CpG islands and shores, while hypomethylation occurred globally across intragenic regions[32]. We observed a strong co-localization of the hypermethylated regions with H3K27me3 marked (polycomb repressed) regions in HMEC cells, supporting the finding that many polycomb-regulated genes are predisposed to aberrant methylation in cancer[33]. We identified 308 genes affected by promoter hypermethylation and functional analysis revealed significant enrichment of genes and transcription factors involved in development and differentiation, as well as DNA binding, homeobox proteins and transcriptional regulation. Hypermethylation of homeobox genes has been previously reported in breast cancer and associated with disease progression and poor patient prognosis[15, 16, 34, 35]. Genes encoding glycoproteins were also enriched in the functional analysis. A significant function of glycoproteins is that of directing immune response[36]. This

is particularly poignant since several gene expression modules associated with immune response have been used to predict TNBC patient outcome[37, 38, 39, 40, 41]. Many of the aberrant cancer promoter hypermethylation events affect genes already silenced in the tissue of origin and therefore considered to be passenger events that do not actively contribute to cancer initiation or progression[42]. To identify potential driver methylation events we highlighted genes that were both down-regulated in TNBC tumours and recurrently mutated in breast cancer. Twelve methylated genes were identified as both mutated and down-regulated, including ROBO3 and SEMA5A that are part of the axon guidance pathway, recently implicated in tumour initiation and progression[27]. In total, promoter hypermethylation affects seven members of the axon guidance pathway. Although the mechanism by which axon guidance drives cancer progression is not completely understood, our data supports a potential causal role for DNA methylation for many of these family members in TNBCs.

Using an independent TNBC cohort from the TCGA data we validated 36 TNBC DMRs comprising 20 genes. Strikingly, 4 of the 20 genes encoded zinc finger proteins (ZNFs). Individual ZNFs and even some clusters of ZNF genes have been found hypermethylated and silenced in several tumour types[43, 44, 45, 46]. In addition, methylation of other ZNF genes have potential prognostic value in prostate and bladder cancer[47, 48]. Although the mechanisms by which aberrant ZNF expression facilitates oncogenesis are not completely understood, ZNFs are included in 2 independently derived, TNBC specific, multi-gene expression classifiers (TN45 and Buck 14)[38, 39].

Recent studies have identified non-TNBC as more heavily methylated compared to TNBC[16]. In our study, we found that a distinct population of both ER+ and ER-ve tumours are associated with extensive methylation across the DMRs, more representative of a CpG island methylator phenotype (CIMP)[29]. Interestingly, a previous report describes the breast-CIMP (B-CIMP) group comprised solely of ER+ve tumours[16], however, our results show that coordinated hypermethylation can also occur in ER-ve disease. We also identified 3 distinct methylation clusters of TNBC tumours based on our DMRs. The largely hypomethylated profile was associated with better survival within the first 5 years post-diagnosis compared to the more heavily methylated subtypes. Interestingly, the medium methylated cluster was associated with the worst survival. Proof of concept that

methylation can be used to stratify breast cancer subtypes was recently demonstrated by TCGA, where DNA methylation data was used to classify breast cancer into 5 distinct subtypes, however, each of the five methylation groups were represented by multiple tumour subtypes and the relationship between methylation and prognosis was not explored[13]. Here, we also identified 17 individual DMRs capable of stratifying TNBC patients into good and poor prognosis groups. Notably, these regions predominately overlap with DNAaseI hypersensitive regions and contain conserved transcription factor binding sites highlighting their potentially significant role in transcriptional regulation. Of the genes listed, many, including WT1, WT1-AS, DMRTA1 and HOXB13, have been previously identified as hypermethylated in numerous cancer subtypes including breast cancer[49, 50, 51, 52]; although associations with patient prognosis were not defined in these studies.

Finally, three “survival” DMRs, span the bi-directional promoter and gene bodies of WT1 gene and its anti-sense counter-part WT1-AS. Wilms tumour protein (WT1) protein is an extensively studied transcription factor essential for normal development of the urogenital system and deregulated across many cancer types[31]. In breast cancer high mRNA levels of WT1 were reported to be associated with poor patient survival[53] and positive modulation of expression of WT1 by its antisense transcript WT1-AS[54, 55]. Our observed patterns of methylation and survival support an extensive body of evidence on the tight epigenetic transcriptional regulation of WT1 and its role in breast cancer prognosis. More specifically, high levels of methylation across regions spanning gene bodies of WT1 and WT1-AS genes correlates with elevated levels of expression and poor survival. Whereas, hypermethylation spanning the bi- directional promoter are associated with decreased WT1 and WT1-AS expression and better survival.

Cumulatively, the work presented here highlights the prognostic potential of DNA methylation in the stratification of TNBC patient prognosis. We identified individual potential prognostic biomarkers of patient outcome as well as providing the first evidence to suggest that DNA methylation potentially could be used to stratify TNBC subtypes that are associated with distinct prognostic profiles. Both observations warrant further clinical investigation in larger independent cohorts as these signatures may in the future provide valuable tools in the management of TNBC.

Methods

Breast cancer tissue samples

Human tissue samples representing normal and tumour breast from fresh frozen and formalin fixed paraffin embedded tissue were obtained for this study. Only samples that were classified as triple negative Grade 3 ductal adenocarcinomas (Supplementary Table 1) were included. The study protocol was approved by the Hunter New England Human Research Ethics Committee (NSW HREC Reference No: HREC/09/HNE/153), Newcastle, New South Wales, Australia and the Princess Alexandra Hospital Human Research Ethics Committee (PAH HREC)(Research Protocol: 2007/165), Brisbane, Queensland.

DNA isolation from formalin-fixed paraffin embedded material

DNA isolation from microdissected formalin fixed paraffin embedded tissue was performed using the Gentra Puregene Genomic DNA purification tissue kit according to the manufacturer's instructions (Qiagen). 5 x 1mm cores or 5 x 10um full faced sections were used for each extraction. The de-paraffinization step was carried out as follows: the paraffin samples was cut into small piece, 500ul Xylene was added and incubated at 55°C for 5 mins, and the tissue was pelleted at 16,000g for 3 mins, discarding the Xylene. After repeating this step, 500ul 100% EtOH was added for 5 mins at room temperature with constant mixing and the tissue collected by centrifugation @16,000g for 3 mins. The EtOH step was repeated and the tissue pellet dried for 10 mins. 300ul of cell lysis solution was added and the tube incubated for 70°C for 10 mins, followed by addition of 20ul Proteinase K (20mg/ml) to each sample and vortexing for 20 secs and incubation in a 55°C block overnight with constant vortexing. The following day a further 10ul proteinase K was added, vortexed for 20 secs and further incubated at 55°C until the samples appear clear. 1ul RNase A solution (100mg/ml) was added, mixed by inverting 25 times and incubated at 37°C for 1 hr. The sample was placed on ice to quickly cool it. Then 100ul protein precipitation solution was added to the cell lysates, vortexed for 20 secs, incubated on ice for 5 mins, and centrifuged at full speed for 5 mins at 4°C to pellet the protein precipitate. The supernatant containing the DNA was carefully removed into a clean microcentrifuge tube. The DNA was precipitated with 300ul 100% isopropanol and 2ul glycogen (20mg/ml)

were added if low yield was expected (<1ug). The solutions were mixed by inversion (50 times) followed by centrifugation for 10 mins at 4°C.

The DNA pellet was washed with 70% EtOH, air-dried and dissolved in 20ul H20. To dissolve the pellet it was incubated for 1 hr at 65°C with constant vortexing.

Enrichment of methylated DNA by MBDCap

The MethylMinerTM Methylated DNA Enrichment Kit (Invitrogen) was used to isolate methylated DNA from 500 ng–1 µg of genomic FFPET DNA and was sonicated to 100–500 bp. MBD-Biotin Protein (3.5 µg) was coupled to 10 µl of Dynabeads M-280 Streptavidin according to the manufacturer's instructions. The MBD-magnetic bead conjugates were washed three times and resuspended in 1 volume of 1x Bind/Wash buffer. The capture reaction was performed by the addition of 500ng - 1 µg sonicated DNA to the MBD-magnetic beads on a rotating mixer for 1 h at room temperature. All capture reactions were done in duplicate. The beads were washed three times with 1x Bind/Wash buffer. The bound methylated DNA was eluted as a single fraction with a single High Salt Elution Buffer (2,000 mM NaCl). Each fraction was concentrated by ethanol precipitation using 1 µl glycogen (20 µg/µl), 1/10 volume of 3 M sodium acetate, pH 5.2 and 2 sample volumes of 100% ethanol and resuspended in 60 µl H2O. Enrichment of methylated DNA after capture was previously assessed by quantitative PCR of control genes of known methylation status; namely EN1 (heavily methylated) and GAPDH (unmethylated)[22].

Preparation of MBDCap-Seq libraries and Illumina sequencing

10ng DNA of MBDCap enriched DNA was prepared for Ilumina sequencing using the Illumina ChIP-Seq DNA sample prep kit (IP-102-1001) according to the manufacturer's instructions. The library preparation was analyzed on Agilent High Sensitivity DNA 1000 Chip. Each sample was sequenced on one lane of the GA11x.

Computational analysis of MBDCap-Seq data

Sequenced reads were aligned to the hg18 version of the human genome with bowtie. Reads with more than three mismatches and reads mapping to multiple positions were removed. Finally, multiple reads mapping to exactly the same genomic coordinate were eliminated and only one read was retained for downstream analysis. Alignment statistics

for samples used in this study are given in Supplementary Table 5. MBDCap-Seq platform was previously shown to interrogate CpG dense regions of the genome[23]. In order to accurately delineate regions of the genome assayable by MBDCap-Seq we used fully methylated sample (SssI blood sample) to guide us to the genomic regions attracting sequenced tags. More specifically, we applied findPeaks peak calling utility from HOMER suite of programs[56] to fully methylated sample (with parameter settings of -style histone -size 300 -minDist 300 -tagThreshold 18) to identify 230,655 regions covering approximately 116Mbp of the genome. We interchangeably refer to these regions as regions of interest or SssI regions. For each MBDCap-Seq sample to be analyzed, we computed the number of sequenced tags overlapping SssI regions, which resulted in table of counts where columns are samples and rows are SssI regions. We used edgeR Bioconductor package[57] (http://www.bioconductor.org/packages/release/bioc/html/edgeR.html) to model distribution of reads between normal (n=6) and tumor (n=19) group of samples in the discovery cohort. Since edgeR package does not support modeling of paired and unpaired data simultaneously we performed two separate analyses, a paired analysis with 6 normal/tumor pairs and unpaired analysis with all the samples, and then intersected the results. We found 822 hyper-methylated and 43 hypo-methylated regions at FDR threshold of 0.05 in both paired and unpaired analyses.

Clustering of MBDCap-Seq data

The number of reads mapping to a particular region of genome depends not just on the average level of methylation in the region, but also on other factors, such as density of methylated CpG nucleotides. In order to compare MBDCap-Seq readout to other more quantitative technologies such as HM450K and Sequenom we used fully methylated MBDCap-Seq sample to normalize MBDCap-Seq readouts for samples in the discovery cohort. More specifically, let be the number of tags overlapping region and be the total number of tags overlapping SssI regions in the sample to be normalized and and be the corresponding numbers in the control sample. Then, the normalized number of tags overlapping the region is given by

$\log (X_i / N.M / Y_i + 1)$

We used normalized tag counts for heatmap visualization in Fig. 1, for comparison to HM450K in Supp. Fig. 2, and for comparison to Sequenom in Supp. Fig. 4

Functional annotations of the genome

CpG island annotation for hg18 was obtained from UCSC genome browser. The location of CpG island shores was derived from CpG islands by taking +/-2Kb flanking regions and removing any overlaps with CpG islands. RefSeq transcript annotation for hg18 was obtained from UCSC genome browser. Promoters were defined as +2000/-100bp around transcription start site (TSS). Intergenic regions were defined as regions complementing transcript regions extended to +/-2Kb around the transcripts. HMEC ChromHMM annotations for hg18 were downloaded from ENCODE. The original annotation partitions the HMEC genome into 15 functional states (see Fig. 1b in 24). In Fig. 1C and Supp. Fig. 1B, for brevity, we collapsed the three original promoter states into one promoter state and the four original enhancer states into one enhancer state.

Enrichment analysis statistical methods

For the enrichment analysis of hypermethylated regions we used hyper-geometric test to assess enrichment of various functional annotations of the genome in the set of differentially methylated regions. For a given functional annotation represented by a set of genomic regions, fraction of SssI regions (regions assayable by MBDCap-Seq) overlapping functional annotation was compared to the fraction of hyper-/hypo- methylated regions overlapping functional annotation using hyper-geometric distribution. For enrichment analysis of genes affected by promoter hyper- methylation, firstly, we used DAVID functional annotation tool[25] to carry analysis against gene sets defined by SP_PIR_KEYWORDS annotation. Secondly, we used hyper-geometric test to assess enrichment of additional gene sets in the set of genes affected by promoter hyper-methylation[26]. In both analyses, the set of 15,643 RefSeq genes with promoters overlapping SssI regions was used as a background.

Sequenom quantitative massARRAY methylation analysis

Sequenom MassARRAY methylation analysis was performed according to Coolen et al.[58]. Briefly, 500ng of FFPET clinical sample and cell line DNA (Supplementary Table 2) was extracted and bisulphite treated using the standard bisulphite protocol[59]. As controls for the methylation analysis, whole genome amplified (WGA) DNA (0% methylated) and M.SssI treated DNA (100% methylated) were bisulphite treated in parallel. The primers were designed using the EpiDesignerBETA soft- ware from Sequenom (see Supplementary Table 6 for sequences). Each reverse primer has a T7-promoter tag (5-CAG TAA TAC GAC TCA CTA TAG GGA GAA GGCT-3) and each forward primer has a 10-mer tag (5-AGG AAG AGA G-3). Upon testing these primers on bisulphite treated DNA, all the primers gave specific PCR products at a Tm of 60°C. In order to check for potential PCR bias towards methylated or non- methylated sequences, we used serological DNA (Millipore) as a 100% methylated control and Whole Genome Amplified human blood DNA as a 0% methylated control. The PCRs were optimized and performed in triplicate using the conditions: 95°C for 2 min, 45 cycles of 95°C for 40 sec, 60°C for 1 min and 72°C for 1 min 30 sec and final extension at 72°C for 5 min. After PCR amplification, the triplicates were pooled and a Shrimp Alkaline Phosphatase (SAP) treatment was performed using 5 µl of the PCR product as template. 2 µl of the SAP-treated PCR product was taken and subjected to in vitro transcription and RNaseA Cleavage for the T-cleavage reaction. The samples were purified by resin treatment and spotted on a 384-well SpectroCHIP by a MassARRAY Nanodispenser. This was followed by spectral acquisition on a MassARRAY Analyser Compact matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry. The results were then analysed by the EpiTYPER software V 1.0 which gives quantitative methylation levels for individual CpG sites. The average methylation ratio was calculated by averaging the ratios obtained from each CpG site. For the Sequenom validation, sample sizes were determined for a two sample t-test with a 2-sided alpha of 0.01, assuming 5 regions were to be investigated. Assuming the difference in average methylation levels is 0.25 (tumors: SD=0.2, normals: SD=0.05), in order to have 90% power to establish a significant difference between tumor and normal samples, 15 samples per group were required. The calculations are based on preliminary data from the lab on methylation levels in breast cancer and normal samples (unpublished).

Acquisition of TCGA data

Throughout the paper we used several molecular datasets from TCGA breast cancer (BRCA) cohort. Clinical annotation of samples was obtained from the marker TCGA BRCA publication (Supplementary Table 1 in [13]; Supplementary Table 7). Raw HM450K methylation data (Level 1) was obtained from TCGA data portal in January 2012. Methylation data spanned 67 normal and 354 tumor ER+ve samples, 16 normal and 105 tumor ER-ve samples, and 9 normal and 73 tumor TNBC samples. Processed array expression data (Level 3) was obtained from TCGA data portal in March 2012. Expression data spanned 52 normal and 406 tumor ER+ve samples, 9 normal and 118 tumor ER-ve samples, and 8 normal and 89 tumor TNBC samples. Processed RNA- Seq expression data (Level 3) was obtained from TCGA data portal in December 2012. Expression data spanned 73 normal and 588 tumor ER+ve samples, 19 normal and 174 tumor ER-ve samples, and 12 normal and 119 tumor TNBC samples. Summary of TCGA BRCA mutation data was obtained from COSMIC database (http://cancer.sanger.ac.uk/cosmic/study/overview?study_id=414). The summary lists mutations in gene coding regions across patients including both synonymous and non-synonymous amino-acid substitutions. We consider a gene as mutated if it appears at least two times in the list (Supplementary Table 8).

Analysis of HM450K methylation data

The raw HM450K data was preprocessed and background normalized with Biconductor minfi package using preprocessIllumina(..., bg.correct = TRUE, normalize = "controls", reference=1) command; resulting M-Values were used for statistical analyses 60 and Beta-Values for heatmap visualizations and clustering. To identify TNBC specific HM450K probes we carried out t-test comparison between TNBC (n=73) and non-TNBC (n=386) tumors. This analysis resulted in 282 probes having adj. p-value less than 0.05 and estimated mean difference of methylation between TNBC and non-TNBC tumors of at least 10%; these probes were declared as TNBC specific. Regions overlapping 3 or more TNBC specific probes were declared as TNBC specific. For TNBC specific signature we trained a Partial Least Squares model as implemented in the c a r e t R package[61, 62] to classify tumours into TNBC and non-TNBC based on methylation values of 282 TNBC specific probes. The tumour samples in the TCGA HM450K cohort were randomly partitioned into equal size training/testing sets. The model parameters were derived from

training set and then applied to make predictions on the testing set. The performance of the model was assessed using test set predictions.

Analysis of expression data

Differential expression analysis between normal (n=8) and tumor (n=89) TNBC samples was carried out with Bioconductor limma package. Since only subset of tumor samples had paired adjacent normal samples, patient data was treated as random effect using limma's duplicateCorrelation(…) function. This analysis resulted in 3,017 down-regulated and 3,407 up-regulated genes with adj. p-value less than 0.05 out of 17,655 genes on the array. In Fig. 1F and 1G we only considered genes with SssI regions in their promoter regions reducing the number of down-regulated, up- regulated and total genes to 2,119, 2,722 and 15,543 respectively. We used log transformed RNA-Seq expression values to highlight relationship between methylation and expression for number of candidate regions in Fig. 2C and Supplementary Fig. 11.

Survival analysis

TNBC tumour samples in TCGA HM450K cohort (n=73) were clustered based on methylation beta-values of 4,987 HM450K probes overlapping the 822 hyper- methylated regions. We applied consensus clustering algorithm[63] as implemented in Bioconductor ConsensusClusterPlus package to the 4,987 X 73 methylation matrix with parameters maxK=4, reps=1000, pItem=0.8, pFeature=0.8, clusterAlg="km", distance="euclidean". We used SVD decomposition to reduce the dimension of the methylation matrix to prior to clustering. We chose the three-cluster configuration for downstream survival analysis. Survival analysis was carried out using Cox proportional hazards model as implemented in R s u r v i v a l package against overall survival data (Supplementary Table 7). Survival analysis of cluster data was carried out with cluster membership as an explanatory variable. The BRCA TNBC cohort consists of 73 patients with HM450K methylation data and 12 events. Survival analysis of individual probes (4,987 probes overlapping 822 hyper-methylated DMRs) was carried out with probe methylation status as explanatory variable (univariate analysis) and age, stage and probe methylation status (multivariate analysis). Methylation status was represented by a binary variable, high (higher that the median beta-value for the probe) and low (smaller or equal to the median beta-value for the

probe). Stage was derived from AJCC stage in the clinical annotation of samples. Due to moderate size the cohort we reduced the number of values of the stage variable to two by collapsing stages I, IA, IB, II, IIA, and IIB into one state and stages III, IIIA, IIIB, IIIC, and IV into one state. This resulted in 190 probes with methylation status statistically significantly (p-value < 0.05 in both univariate and multivariate analyses) associated with overall survival in TCGA TNBC patients. Regional aggregation of survival probes identified 17 hyper- methylated DMRs overlapping three or more survival probes. Fourteen regions were associated with poor prognosis, these regions overlapped probes for which high methylation corresponded to lower probability of survival, and three regions were associated with good prognosis.

Summary of other statistical methods

For the enrichment analysis of hypermethylated regions we used hyper-geometric test to assess enrichment of various functional annotations of the genome in the set of differentially methylated regions. For a given functional annotation represented by a set of genomic regions, fraction of SssI regions (regions assayable by MBDCap-Seq) overlapping functional annotation was compared to the fraction of hyper-/hypo- methylated regions overlapping functional annotation using hyper-geometric distribution.

For enrichment analysis of genes affected by promoter hyper-methylation, firstly, we used DAVID functional annotation tool[25] to carry analysis against gene sets defined by SP_PIR_KEYWORDS annotation. Secondly, we used hyper-geometric test to assess enrichment of additional gene sets in the set of genes affected by promoter hyper-methylation[26]. In both analyses, the set of 15,643 RefSeq genes with promoters overlapping SssI regions was used as a background.

Competing Interests

The authors declare that they have no competing interests.

Authors Contributions

Conception and design: S.J.C., M.T., A.D., J.F., R.S., M.A.B., G.F. Development of methodology: C.S., J.S.

Acquisition of data: C.S., E.Z., J.S., S.N., W.L.

Provided clinical samples (preparation of DNA): K.A.K., K.P., W.Q., S.S., K.P., J.D.F.
Analysis and interpretation of the data (eg. Statistical analysis, biostatistics, computational analysis): N.A., M.R., W.L., E.Z., A.S.
Writing, figures, review of the manuscript: E.Z., W.L., A.S., C.S., S.J.C. Study supervision: S.J.C., C.S

Acknowledgments
We thank the Ramaciotti Centre, University of NSW (Sydney, Australia) for genome sequencing. This work is supported by the National Breast Cancer Foundation (NBCF) program and project grants and National Health and Medical Research Council (NHMRC1029579) project grant and NHMRC Fellowship to S.J.C. M.A.B is supported by Cancer Council Queensland and University of Queensland. J.D.F is supported by a fellowship from the National Breast Cancer Foundation (NBCF) Australia.

Accession Codes

Methylation sequence data have been deposited in GenBank/EMBL/DDBJ under the accession code GSE58020.

References

1. Perou CM, et al. Molecular portraits of human breast tumours. Nature 406, 747-752 (2000).
2. Perou CM. Molecular stratification of triple-negative breast cancers. Oncologist 16 Suppl 1, 61-70 (2011).
3. Blows FM, et al. Subtyping of breast cancer by immunohistochemistry to investigate a relationship between subtype and short and long term survival: a collaborative analysis of data for 10,159 cases from 12 studies. PLoS Med 7, e1000279 (2010).
4. Dent R, et al. Triple-Negative Breast Cancer: Clinical Features and Patterns of Recurrence. Clinical Cancer Research 13, 4429-4434 (2007).
5. Jatoi I, Anderson WF, Jeong J-H, Redmond CK. Breast Cancer Adjuvant Therapy: Time to Consider Its Time-Dependent Effects. Journal of Clinical Oncology 29, 2301-2304 (2011).

6. Park YH, et al. Clinical relevance of TNM staging system according to breast cancer subtypes. Annals of Oncology 22, 1554-1560 (2011).
7. Reis-Filho JS, Pusztai L. Gene expression profiling in breast cancer: classification, prognostication, and prediction. The Lancet 378, 1812-1823.
8. Laird PW. The power and the promise of DNA methylation markers. Nat Rev Cancer 3, 253-266 (2003).
9. Szyf M. DNA methylation signatures for breast cancer classification and prognosis. Genome Medicine 4, 26 (2012).
10. Xu Z, Bolick SCE, DeRoo LA, Weinberg CR, Sandler DP, Taylor JA. Epigenome-wide Association Study of Breast Cancer Using Prospectively Collected Sister Study Samples. Journal of the National Cancer Institute 105, 694-700 (2013).
11. Chimonidou M, et al. CST6 promoter methylation in circulating cell-free DNA of breast cancer patients. Clin Biochem 46, 235-240 (2013).
12. Snell C, Krypuy M, Wong EM, Loughrey MB, Dobrovic A. BRCA1 promoter methylation in peripheral blood DNA of mutation negative familial breast cancer patients with a BRCA1 tumour phenotype. Breast Cancer Res 10, 12 (2008).

13. TCGA. Comprehensive molecular portraits of human breast tumours. Nature 490, 61-70 (2012).
14. Holm K, et al. Molecular subtypes of breast cancer are associated with characteristic DNA methylation patterns. Breast Cancer Research 12, R36 (2010).
15. Fackler MJ, et al. Genome-wide methylation analysis identifies genes specific to breast cancer hormone receptor status and risk of recurrence. Cancer Res 71, 6195-6207 (2011).
16. Fang F, et al. Breast cancer methylomes establish an epigenomic foundation for metastasis. Sci Transl Med 3, 75ra25 (2011).
17. Cho YH, et al. Prognostic significance of gene-specific promoter hypermethylation in breast cancer patients. Breast Cancer Res Treat 131, 197- 205 (2012).
18. Stone A, et al. BCL-2 Hypermethylation Is a Potential Biomarker of Sensitivity to Antimitotic Chemotherapy in Endocrine-Resistant Breast Cancer. Molecular Cancer Therapeutics 12, 1874-1885 (2013).

19. Stefansson OA, Villanueva A, Vidal A, Marti L, Esteller M. BRCA1 epigenetic inactivation predicts sensitivity to platinum-based chemotherapy in breast and ovarian cancer. Epigenetics 7, 1225-1229 (2012).
20. Serre D, Lee BH, Ting AH. MBD-isolated Genome Sequencing provides a high-throughput and comprehensive survey of DNA methylation in the human genome. Nucleic Acids Res 38, 391-399 (2010).
21. Rauch T, Pfeifer GP. Methylated-CpG island recovery assay: a new technique for the rapid detection of methylated-CpG islands in cancer. Lab Invest 85, 1172-1180 (2005).
22. Nair SS, et al. Comparison of methyl-DNA immunoprecipitation (MeDIP) and methyl-CpG binding domain (MBD) protein capture for genome-wide DNA methylation analysis reveal CpG sequence coverage bias. Epigenetics : official journal of the DNA Methylation Society 6, 34-44 (2011).

23. Robinson MD, et al. Evaluation of affinity-based genome-wide DNA methylation data: effects of CpG density, amplification bias, and copy number variation. Genome Res 20, 1719-1729 (2010).
24. Ernst J, et al. Mapping and analysis of chromatin state dynamics in nine human cell types. Nature 473, 43-49 (2011).
25. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4, 44-57 (2009).
26. Merico D, Isserlin R, Stueker O, Emili A, Bader GD. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. PLoS One 5, e13984 (2010).
27. Mehlen P, Delloye-Bourgeois C, Chedotal A. Novel roles for Slits and netrins: axon guidance cues as anticancer targets? Nat Rev Cancer 11, 188-197 (2011).
28. Neufeld G, Kessler O. The semaphorins: versatile regulators of tumour progression and tumour angiogenesis. Nat Rev Cancer 8, 632-645 (2008).
29. Hughes LAE, et al. The CpG island methylator phenotype: what's in a name? Cancer Research, (2013).
30. Matys V, et al. TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. Nucleic Acids Res 34, D108-110 (2006).
31. Yang L, Han Y, Suarez Saiz F, Minden MD. A tumor suppressor and oncogene: the WT1 story. Leukemia 21, 868-876 (2007).

32. Shen H, Laird Peter W. Interplay between the Cancer Genome and Epigenome. Cell 153, 38-55 (2013).
33. Cedar H, Bergman Y. Linking DNA methylation and histone modification: patterns and paradigms. Nat Rev Genet 10, 295-304 (2009).
34. Tommasi S, Karm DL, Wu X, Yen Y, Pfeifer GP. Methylation of homeobox genes is a frequent and early epigenetic event in breast cancer. Breast Cancer Res 11, R14 (2009).
35. Pilato B, Pinto R, De Summa S, Lambo R, Paradiso A, Tommasi S. HOX gene methylation status analysis in patients with hereditary breast cancer. J Hum Genet 58, 51-53 (2013).
36. Rudd PM, Elliott T, Cresswell P, Wilson IA, Dwek RA. Glycosylation and the Immune System. Science 291, 2370-2376 (2001).
37. Teschendorff A, Miremadi A, Pinder S, Ellis I, Caldas C. An immune response gene expression module identifies a good prognosis subtype in estrogen receptor negative breast cancer. Genome Biology 8, R157 (2007).
38. Yau C, Esserman L, Moore D, Waldman F, Sninsky J, Benz C. A multigene predictor of metastatic outcome in early stage hormone receptor-negative and triple-negative breast cancer. Breast Cancer Research 12, R85 (2010).
39. Kuo WH, et al. Molecular characteristics and metastasis predictor genes of triple-negative breast cancer: a clinical study of triple-negative breast carcinomas. PLoS One 7, e45831 (2012).
40. Rody A, et al. A clinically relevant gene signature in triple negative and basal- like breast cancer. Breast Cancer Research 13, R97 (2011).
41. Hallett RM, Dvorkin-Gheva A, Bane A, Hassell JA. A Gene Signature for Predicting Outcome in Patients with Basal-like Breast Cancer. Sci Rep 2, (2012).
42. Sproul D, et al. Transcriptionally repressed genes become aberrantly methylated and distinguish tumors of different lineages in breast cancer. Proc Natl Acad Sci U S A 108, 4364-4369 (2011).
43. Cheng Y, et al. KRAB Zinc Finger Protein ZNF382 Is a Proapoptotic Tumor Suppressor That Represses Multiple Oncogenes and Is Commonly Silenced in Multiple Carcinomas. Cancer Research 70, 6516-6526 (2010).
44. Lleras RA, et al. Hypermethylation of a Cluster of Krüppel-Type Zinc Finger Protein Genes on Chromosome 19q13 in Oropharyngeal Squamous Cell Carcinoma. The American Journal of Pathology 178, 1965-1974 (2011).

45. Huang R-L, et al. Methylomic Analysis Identifies Frequent DNA Methylation of Zinc Finger Protein 582 (ZNF582 in Cervical Neoplasms. PLoS ONE 7, e41060 (2012).

46. Severson PL, Tokar EJ, Vrba L, Waalkes MP, Futscher BW. Coordinate H3K9 and DNA methylation silencing of ZNFs in toxicant-induced malignant transformation. Epigenetics 8, 1080-1088 (2013).

47. Vanaja DK, Cheville JC, Iturria SJ, Young CYF. Transcriptional Silencing of Zinc Finger Protein 185 Identified by Expression Profiling Is Associated with Prostate Cancer Progression. Cancer Research 63, 3877-3882 (2003).

48. Reinert T, Borre M, Christiansen A, Hermann GG, Orntoft TF, Dyrskjot L. Diagnosis of bladder cancer recurrence based on urinary levels of EOMES, HOXA9, POU4F2, TWIST1, VIM, and ZNF154 hypermethylation. PLoS One 7, e46297 (2012).

49. Rodriguez BA, et al. Epigenetic repression of the estrogen-regulated Homeobox B13 gene in breast cancer. Carcinogenesis 29, 1459-1465 (2008).

50. Bruno P, et al. WT1 CpG islands methylation in human lung cancer: a pilot study. Biochem Biophys Res Commun 426, 306-309 (2012).

51. Ghoshal K, et al. HOXB13, a Target of DNMT3B, Is Methylated at an Upstream CpG Island, and Functions as a Tumor Suppressor in Primary Colorectal Tumors. PLoS ONE 5, e10338 (2010).

52. Okuda H, et al. Epigenetic inactivation of the candidate tumor suppressor gene HOXB13 in human renal cell carcinoma. Oncogene 25, 1733-1742 (2006).

53. Miyoshi Y, et al. High Expression of Wilms' Tumor Suppressor Gene Predicts Poor Prognosis in Breast Cancer Patients. Clinical Cancer Research 8, 1167- 1171 (2002).

54. Moorwood K, Charles AK, Salpekar A, Wallace JI, Brown KW, Malik K. Antisense WT1 transcription parallels sense mRNA and protein expression in fetal kidney and can elevate protein levels in vitro. J Pathol 185, 352-359 (1998).

55. Dallosso AR, et al. Alternately spliced WT1 antisense transcripts interact with WT1 sense RNA and show epigenetic and splicing defects in cancer. Rna 13, 2287-2299 (2007).

56. Heinz S, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell 38, 576-589 (2010).

57. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139-140 (2010).
58. Coolen MW, Statham AL, Gardiner-Garden M, Clark SJ. Genomic profiling of CpG methylation and allelic specificity using quantitative high-throughput mass spectrometry: critical evaluation and improvements. Nucleic Acids Res 35, e119 (2007).
59. Clark SJ, Harrison J, Paul CL, Frommer M. High sensitivity mapping of methylated cytosines. Nucleic Acids Res 22, 2990-2997 (1994).
60. Du P, et al. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. BMC Bioinformatics 11, 587 (2010).
61. Kuhn M. Building Predictive Models in R Using the caret Package. Journal of Statistical Software 28, (2008).
62. Mevik B-H, Wehrens R. The pls Package: Principal Component and Partial Least Squares Regression in R. Journal of Statistical Software 18, (2007).
63. Tamayo SMP, Mesirov J, Golub T. Consensus Clustering: A Resampling- Based Method for Class Discovery and Visualization of Gene Expression Microarray Data. Machine Learning 52, 91-118 (2003).

Figure Legends

Figure 1. MBDCap-seq identifies DMRs in discovery cohort. A heatmap showing methylation profile of 822 hypermethylated (A) and 43 hypomethylated regions (B) across a cohort of 19 tumour and 6 matched normal samples in the discovery cohort. Columns are samples and rows are regions. The level of methylation (number of reads normalized with respect to fully methylated sample) is represented by a colour scale – blue for low levels and red for high levels of methylation. (C) A bar plot showing association of DMRs across functional/regulatory regions of the genome – (i) CpG islands and shores, (ii) RefSeq transcripts, and (iii) Broad ChromHMM HMEC annotation. The height of the bars represents the level of enrichment measured as a ratio between the frequency of hypermethylated (pink) or hypomethylated (blue) regions overlapping a functional element over the expected frequency if such overlaps were to occur at random in the genome. Statistically significant enrichments (p-value < 0.05; hyper-geometric test) are marked with an asterisk. (D) Sequenom validation of five hypermethylated regions -- FERD3L, C9orf125, HMX2, NPY and SATB2 -- is shown for an independent cohort of TNBC

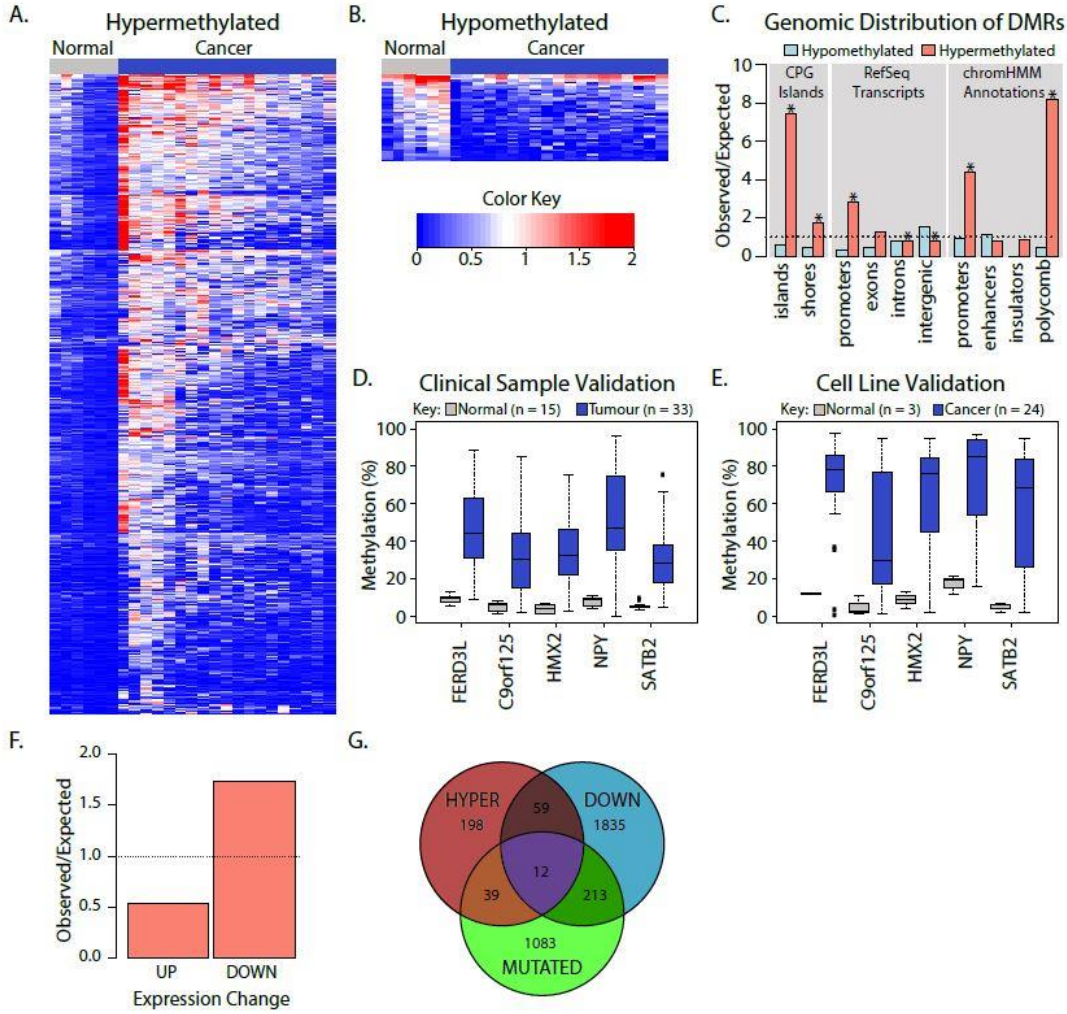
samples (normal=15; tumour =33) and (E) a panel of breast cancer cell lines (normal=3; cancer=24). For each region boxplots displaying the distribution of methylation levels are shown in grey/blue for normal/tumour samples/cell lines. (F) A bar plot showing enrichment of genes with promoter hypermethylation in sets of genes that are up-/down-regulated in the TCGA cohort of TNBC tumour as compared to matched normal samples. The height of the bars represents the level of enrichment measured as a ratio between the observed number of up-/down-regulated genes with promoter hypermethylation to the expected number of such genes. (G) A Venn diagram showing overlap between genes with promoter hypermethylation, genes down-regulated in TCGA TNBC cohort (hyper-geometric test; FC 1.73; p-value <<0.001) and genes with two or more mutation (hyper-geometric test; FC 1.92; p-value <<0.001) in TCGA breast cancer cohort.

Figure 2. Methylation profile of candidate DMRs in the TCGA breast cancer cohort. (A) A heatmap showing methylation profile of TCGA breast cancer samples across 4,987 HM450K probes overlapping hypermethylated DMRs identified in the discovery cohort. Rows are probes and columns are TCGA breast cancer samples profiled on HM450K – 83 normal, 105 ER-VE tumour, and 354 ER+VE tumour samples. (B) A classifier (Partial Least Squares model) based on methylation values of 282 TNBC-specific probes assigns TCGA HM450K tumour samples into TNBC and non-TNBC with high accuracy; ROC analysis yields AUC of 0.9; assigning samples to highest scoring class (TNBC or non-TNBC) yields sensitivity of 0.72 and specificity of 0.94. The TCGA HM450K cohort was randomly split into training set (TNBC n=37; non-TNBC n=193) and testing set (TNBC n=36; non-TNBC n-193). The model was trained on training set and prediction accuracy assessed on testing set.(C) Box plots showing distribution of methylation levels for two adjacent regions on chromosome 19 in TCGA normal (n=83), TNBC tumour (n=73), and other breast tumour samples (n=354). These two regions which span the promoters of ZNF154 and the adjacent ZNF671 gene, are hypermethylated in the discovery cohort and exhibit regional TNBC specific hypermethylation in TCGA cohort, i.e. they are more heavily methylated in TNBC tumours as compared to normal and other tumour subtypes, as shown in the box plots (t-test; mean diff. > 0.1; p-value < 0.05). (D) Box plots showing distribution of expression levels of ZNF154 and ZNF671 genes in TCGA normal (n=92), TNBC tumour (n=119), and ER+VE tumour (n=588) samples. The difference in expression

of TNBC tumours versus ER+VE is found significant for both genes (t-test; ZNF154 mean diff. -0.51 with p-value 9.04e-10; ZNF671 mean diff. -0.85 and p-value of 5.50e-21).

Figure 3. Methylation profile stratifies TNBC tumours into survival subgroups
(A) Unsupervised clustering with 4,987 HM450K probes overlapping 822 hypermethylated DMRs identified in the discovery cohort separates TCGA TNBC tumours (n=73) into three main clusters. The heatmap shows the methylation profile of TCGA TNBC tumours and cluster dendrogram. The three clusters are color-coded with the red cluster exhibiting the highest methylation (TNBC.high), the blue cluster exhibiting the lowest methylation (TNBC.low) and the orange cluster exhibiting an intermediate level of methylation (TNBC.medium). (B) A Kaplan-Meier plot showing survival curves for the patients in the three clusters defined in (A). Additionally, individual regions of hypermethylation in the discovery cohort overlap with survival associated probes in the TCGA cohort; including (C) intergenic loci, (D) intragenic loci (e.g. the HOXB13 gene body) and (E) promoter associate loci (e.g. ZNF254 promoter). (F) Association with survival for three adjacent regions -- chr11-11623, chr11-4047, and chr11-1210 -- spanning the WT1/WT1-AS locus is shown. These three regions are hypermethylated in the discovery cohort and overlap several probes showing statistically significant association with overall survival in both univariate and multivariate analyses. For each region the methylation profile of TCGA TNBC tumour (n=73) and adjacent normal samples (n=9) across overlapping survival probes is shown as a heatmap. The Kaplan-Meyer plots for each of the overlapping survival probes is shown as well with corresponding hazard ratios and p-values from Cox proportional hazards model; values in parentheses correspond to multivariate analysis.

Figure 1


A.
Hypermethylated
Normal
Cancer
B.
Hypomethylated
Normal
Cancer
Color Key
0
0.5
1
1.5
2
C.
Genomic Distribution of DMRs
Hypomethylated
Hypermethylated
CPG Islands
RefSeq Transcripts
chromHMM Annotations
Observed/Expected
islands
shores
promoters
exons
introns
intergenic
promoters
enhancers
insulators
polycomb
D.
Clinical Sample Validation
Key: Normal (n = 15) Tumour (n = 33)
Methylation (%)
FERD3L
C9orf125
HMX2
NPY
SATB2
E.
Cell Line Validation
Key: Normal (n = 3) Cancer (n = 24)
Methylation (%)
FERD3L
C9orf125
HMX2
NPY
SATB2
F.
Observed/Expected
UP
DOWN
Expression Change
G.
HYPER
198
59
DOWN
1835
12
39
213
1083
MUTATED

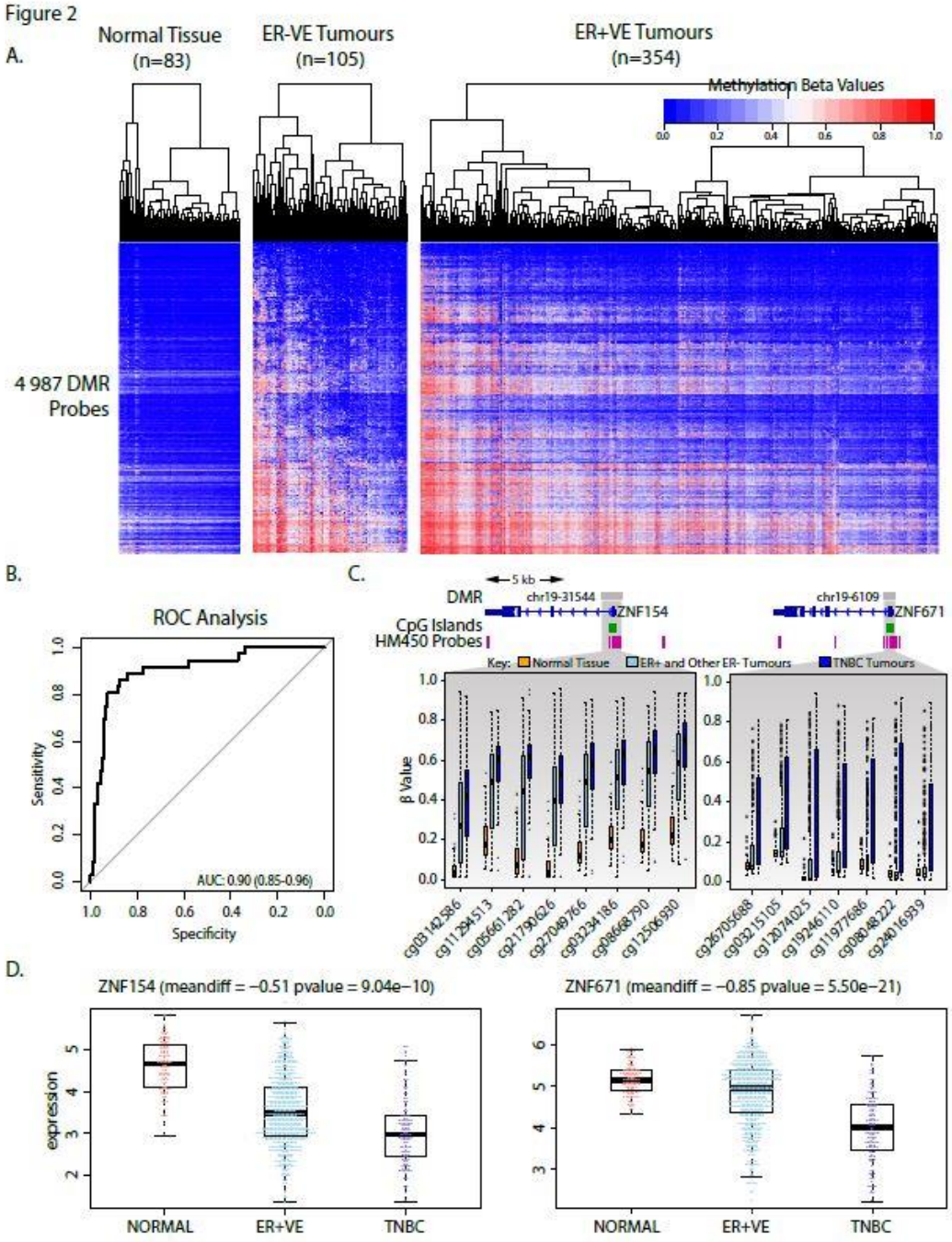
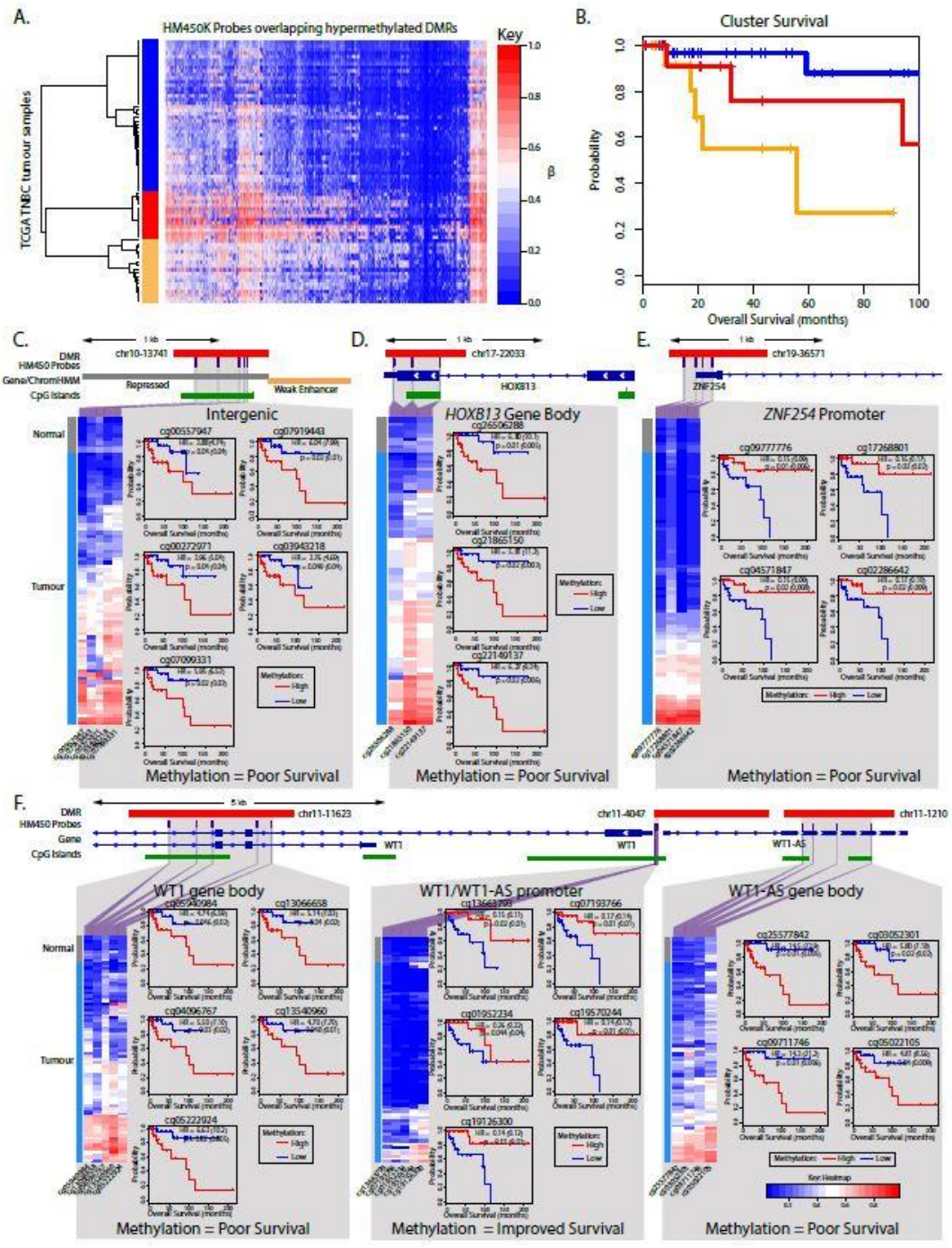


Figure 2
A.
Normal Tissue (n=83)
ER-VE Tumours (n=105)
ER+VE Tumours (n=354)
Methylation Beta Values
4 987 DMR Probes
B.
ROC Analysis
Sensitivity
Specificity
AUC: 0.90 (0.85-0.96)
C.
5 kb
DMR
chr19-31544
ZNF154
chr19-6109
ZNF671
CpG Islands
HM450 Probes
Key: Normal Tissue ER+ and Other ER- Tumours TNBC Tumours
β Value
cg08142586
cg11294513
cg05661282
cg21790626
cg27049766
cg03234186
cg08668790
cg12506930
cg26705688
cg03215105
cg12074025
cg19246110
cg11977686
cg08048222
cg24016939
D.
ZNF154 (meandiff = −0.51 pvalue = 9.04e−10)
ZNF671 (meandiff = −0.85 pvalue = 5.50e−21)
expression
NORMAL
ER+VE
TNBC

Figure 3


A.
HM450K Probes overlapping hypermethylated DMRs
Key
TCGA TNBC tumour samples
β
B.
Cluster Survival
Probability
Overall Survival (months)
C.
DMR
HM450 Probes
Gene/ChromHMM
CpG Islands
chr10-13741
Repressed
Weak Enhancer
Intergenic
Normal
Tumour
Methylation = Poor Survival
D.
chr17-22033
HOXB13
HOXB13 Gene Body
Methylation = Poor Survival
E.
chr19-36571
ZNF254
ZNF254 Promoter
Methylation = Poor Survival
F.
chr11-11623
chr11-4047
chr11-1210
WT1
WT1-AS
WT1 gene body
WT1/WT1-AS promoter
WT1-AS gene body
Methylation = Poor Survival
Methylation = Improved Survival
Methylation = Poor Survival

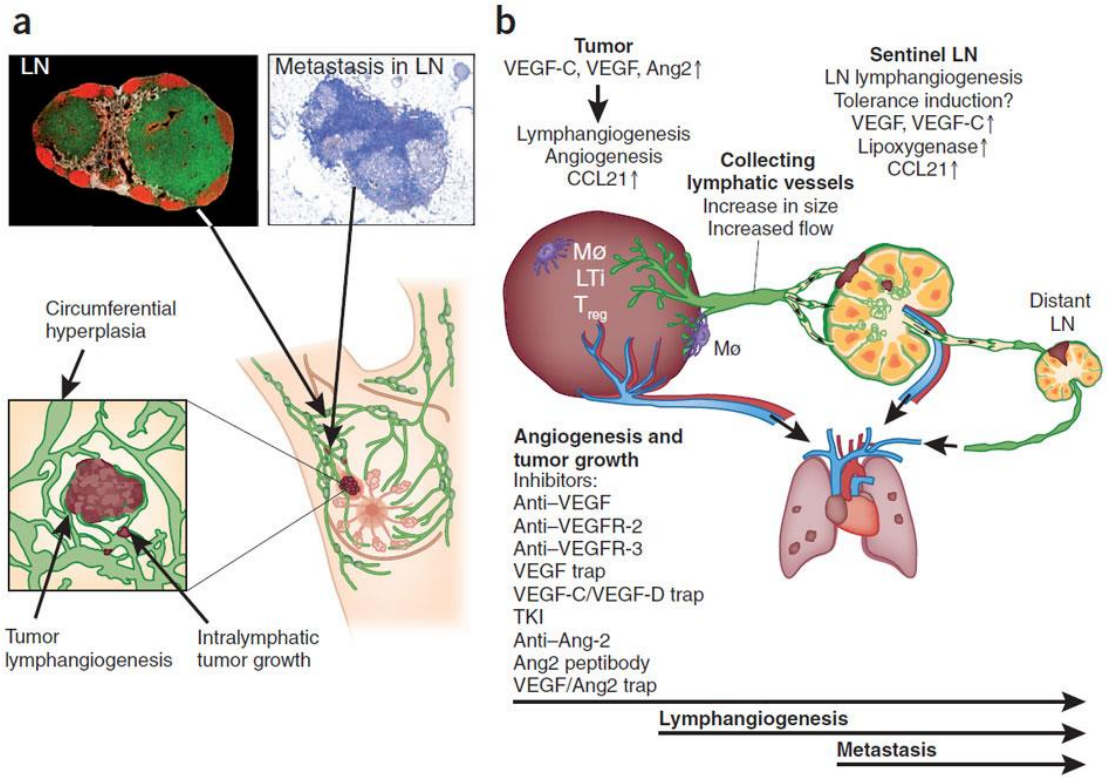
Chapter 10

Discussion

10.1. General Discussion

Breast cancer is one of the commonest cancers in the developed world. In Australia there are more than 14,000 new cases per year with more than 2,500 deaths each year. Breast cancer is the commonest cancer in women and whilst the overall survival is approximately 85% it is still a leading cause of cancer mortality in women. Despite extensive funding for breast cancer relatively little progress has been made on the introduction of new predictive or prognostic biomarkers. The standard clinicopathological biomarkers include tumour size, grade and lymph node status, oestrogen receptor, progesterone receptor and HER2 ISH. In some locations Ki-67 is also included as a biomarker to distinguish tumours that have a higher proliferative index that are more likely to respond to chemotherapy, but this is not uniformly included as a prognostic marker across all laboratories in Australia. Different tumour types are associated with different patterns of metastases and even within the same tumour location there are different patterns of spread. Breast cancer, prostate cancer, papillary thyroid cancer, gastric carcinoma and melanoma have a relatively high frequency of lymph node metastases in contrast to renal cell carcinomas and soft tissue tumours which disseminate by haematogenous metastasis. This variation in the pattern of dissemination in different tumour types does suggest that there are inherent differences in the biological make-up of the different tumours and that haematogenous spread is not contingent on prior lymph node metastases. Lymphangiogenesis and mechanisms of lymphatic spread by tumours have been extensively investigated and some of the pathways have been elucidated (Alitalo, 2011) (Figure 8).

Figure 8: “Mechanisms contributing to lymphatic metastasis. (a) Top left, the normal lymph node has been stained with antibodies detecting B cells (red), T cells (green) and lymphatic sinusoid (white). Top right, the eosin-stained (blue) LN shows metastatic foci in gray. Bottom, increased VEGF-C and D secretion by tumour cells or tumour associated inflammatory cells induces tumour lymphangiogenesis, hyperplasia of collecting lymphatic vessels and tumour cell translocation into lymphatic vessels and lymph nodes. (b) Aberrant expression of chemokine receptors such as CCR7 may increase tumour cell migration toward lymphatic vessels”. Modified from Alitalo (Alitalo, 2011).



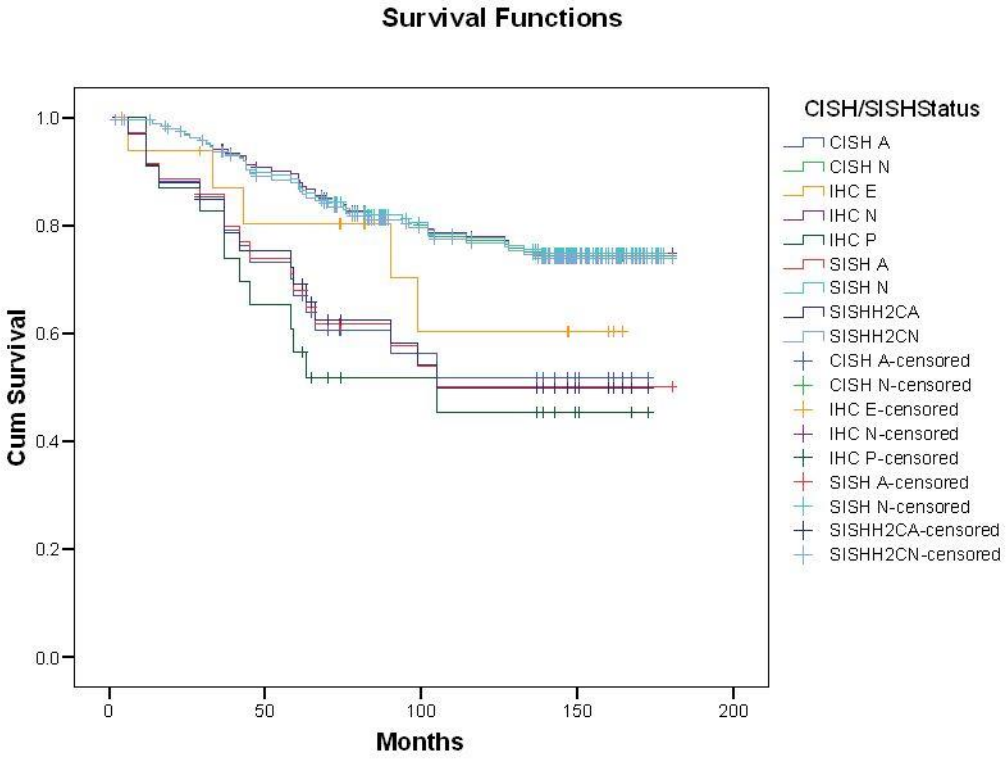
With the development of molecular techniques, multiple potential biomarkers have been identified that associate with lymph node status or with prognosis and patient survival. A limited number of new commercial molecular or protein expression tests have been introduced across the world. These include Oncotype Dx™, Mammaprint™, Endopredict™, Mammostrat™ and PAM50™ (Gyorffy et al., 2015). These tests are based on RT-PCR or gene expression arrays and are expensive (Oncotype Dx™ ~$4500, Mammaprint™ ~$4500-$10,000 and Endopredict™ ~$3000). The uptake of these tests is relatively small in Australia with an estimated uptake of less than 300 tests per year. The tests are also usually restricted to specific subtypes of breast carcinoma such as lymph node negative hormone receptor positive tumours and stratify patients into those patients that will benefit from chemotherapy and those where the benefit from chemotherapy is minimal. Clinical trials have been commenced to determine the utility of Oncotype Dx™ and Mammaprint™ and the cost benefit of these new assays is still uncertain (Bonastre et al., 2014). Even with the development of the molecular assays, lymph node status still remains an independent prognostic variable for breast cancers patients. Whilst the development of sentinel lymph node biopsy has reduced the need for axillary lymph node dissection, the procedure is still associated with morbidity, although at a lower level: “lymphedema (3.5% vs. 19.1%), impaired shoulder range of motion (3.5% vs. 11.3%),

shoulder/arm pain (8.1% vs. 21.1%), and numbness (10.9% vs. 37.7%)" (Langer et al., 2007).

This thesis was aimed at predicting lymph node status from characteristics of the primary tumour using an artificial neural network. To be clinically useful for predicting lymph node status, a test must be able to accurately predict positive nodal status in more than 95% of patients. An assay with an accuracy below this threshold would have an adverse impact on patient outcome due to the prognostic significance of lymph node status in patient survival. The comparator is sentinel lymph node biopsy, and sentinel lymph node biopsy is able to reliably predict axillary lymph node status in 98% of all patients and 95% of those who are node-positive (Cody, 1999).

A breast cancer database was constructed to facilitate the analysis with multiple variables being included in different models. Tissue microarrays were utilised for evaluation of multiple expression markers using immunohistochemistry. The usefulness and the ability of TMAs to accurately reflect the characteristics of the primary tumour has been demonstrated, but the results from TMAs must also be clinically useful and reflect the patient outcome. To confirm that the patient database was an accurate reflection of sporadic breast cancers and that TMA immunohistochemistry correlated with clinical characteristics and patient outcome, HER2 was used as a model. HER2 IHC, HER2 Chromogenic in-situ hybridization (CISH) and HER2 Silver in-situ hybridization (SISH) were performed on TMAs from 230 patients (Chapter 3) (Francis et al., 2009a). The samples used in this analysis covered a range of tumour types, with 87% being infiltrating ductal carcinoma. Of these patient samples, 42% were lymph node negative, 34% were lymph node positive and in the remainder the nodal status was unknown. The median follow up was 139.3 months. Clinical and pathological features represented that seen in sporadic breast cancers. In this study, the two bright field ISH methods showed similar Kaplan-Meier survival curves. Whilst the groups of patients were not corrected for hormone receptor status and treatment, the curves showed that patients with HER2 gene amplified tumours obtained from a TMA analysis, had a worse outcome compared to patients with HER2 negative tumours (Figure 9).

Figure 9: Kaplan-Meier plots for breast cancer specific survival.
Codes: CISH A= low and high amplified cases; CISH N=diploid, polysomy and equivocal cases; SISH A= low and high amplified cases; SISH N=diploid, polysomy and equivocal cases;SISHH2CA= low and high amplified cases utilizing the HER2/CHR17 ratio; SISHH2CN= diploid, polysomy and equivocal cases utilizing the HER2/CHR17 ratio;IHC E= 2+ IHC staining, IHC N= 0 & 1+ IHC staining and IHC P= 3+ IHC staining (Francis et al., 2009a).



In a separate study, androgen receptor (AR) status was evaluated in 73 grade 3 infiltrating ductal carcinomas with positive lymph nodes (Chapter 6) (Peters et al., 2012). Duplicate TMAs were used for analysis and AR expression was detected in 56% (n=41) of the primary breast tumours. 55.5% of the tumours were ER positive, 40% were PR positive and 21.7% were HER2 positive. The lower percentage of ER and PR positive tumours and the higher percentage of HER2 negative tumours is due to section of the tumours for positive lymph node status. There was an association between AR expression and ER, PR and HER2 expression. “The majority of AR-negative tumours were also negative for ER (72%), PR (87%) and HER2 86.6%)” (Peters et al., 2012). In this study, AR expression was identified as a significant prognostic factor for overall patient survival with the 10-year survival of patients with AR positive tumours being 52% versus 22% for patients with AR negative tumours (Peters et al., 2012). DNA methylation, somatic mutations in the AR promoter region and miRNA was also evaluated in this study but did not appear to correlate with AR expression in the breast cancer samples (Peters et al., 2012).

A variety of cell adhesion molecules have also been evaluated in breast cancer. This has included ECM proteins and insulin-like growth factors. TMAs have been used to evaluate the prognostic utility of insulin-like growth factor receptors and extracellular matrix (ECM) interaction-induced proteins in breast carcinoma (Chapter 8) (Plant et al., 2014). Multiple biomarkers of cell adhesion and extracellular matrix interaction were evaluated using immunohistochemistry on TMAs (Table 5) (Plant et al., 2014). The IHC markers were scored and evaluated in a semiquantitative fashion.

Table 5: Product and supplier details for the antibodies used in the study. IgG: immunoglobulin G; mAb: monoclonal antibody; pAb: polyclonal antibody, and; N/A: not applicable (Plant et al., 2014).

Antibody target	Supplier	Supplier details	Product Number	Isotype	Source	Clone
IGF-IRβ	Santa Cruz Biotechnology®	California, USA	sc-713	IgG	Rabbit pAb	C-20
IGF-IIR	Santa Cruz Biotechnology®	California, USA	sc-25462	IgG	Rabbit pAb	H-300
IGFBP-5	R&D Systems®	NSW, Australia	MAB875	IgG_{2b}	Mouse mAb	164503
VN	Epitomics®	California, USA	1730-1	IgG	Rabbit mAb	EP781Y
FN	Novocastra	NSW, Australia	NCL-FIB	IgG_1	Mouse mAb	568
α_v integrin	Calbiochem®	New Jersey, USA	407286	IgG_1	Mouse mAb	272-17E6
β_1 integrin	Abcam®	Massachusetts, USA	ab3167	IgG_1	Mouse mAb	4B7R
Total-AKT1	Cell Signalling Technology®	Massachusetts, USA	2967	IgG_1	Mouse mAb	2H10
P-AKT (Ser473)	Cell Signalling Technology®	Massachusetts, USA	4051L	IgG_{2b}	Mouse mAb	587F11
ERK1/2	Cell Signalling Technology®	Massachusetts, USA	4695	IgG	Rabbit pAb	137F5
P-ERK1/2 (Thr202/Thr204)	Cell Signalling Technology®	Massachusetts, USA	9106	IgG_1	Mouse mAb	E10
CLDN-1	Invitrogen™	VIC, Australia	18-7362	IgG	Rabbit pAb	N/A
SFN	Abcam®	Massachusetts, USA	110-02810	IgG_1	Mouse mAb	3C3
SHARP-2	Sigma-Aldrich®	NSW, Australia	S8443	IgG	Rabbit pAb	N/A
ER	Ventana Medical Systems®	Arizona, USA	790-4325	IgG	Rabbit mAb	SP1
PR	Ventana Medical Systems®	Arizona, USA	790-4296	IgG	Rabbit mAb	1E2
HER2	Ventana Medical Systems®	Arizona, USA	790-2991	IgG	Rabbit mAb	4B5

This study evaluated the cellular localisation of the IHC products within the tumours and identified differences in the subcellular localisation of a range of cell adhesion molecules. Differences were identified between the localisation of the staining pattern between

primary tumours and lymph node metastases, however, markers to predict lymph node status were not identified (Plant et al., 2014).

Extrapolating from all of these studies and general literature, it can be expected that results from TMAs accurately reflect both the results from whole tumour sections and correlate with clinical characteristics within a number of tumour types including breast cancer. TMAs should therefore be able to be utilised to predict lymph node status if there is an association between the biomarker and lymph node status.

Immunohistochemistry was performed on whole tissue sections and TMA sections of a sporadic population of breast carcinomas derived from a large database and these were used for input into an ANN (Appendix 13.5). Multiple models were run to attempt to enable accurate prediction of lymph node status, however despite some models being developed that showed good correlation with lymph node status, these models failed to validate in a separate cohort of breast cancer samples. In comparison, models were able to be developed that accurately predicted breast cancer specific survival over time in the same patient cohort (Appendix 13.6). This suggests that the parameters used in models for prediction of survival are correlated with survival, but those used for lymph node prediction are not. miRNA changes are also able to predict prognosis in breast carcinoma (Chapter 4) (Wee et al., 2012) using patients derived from the same database. A separate study using DNA methylation was also able to predict differences between TNBC patients with different DNA methylation profiles, but a model for prediction of lymph node status was not able to be developed (Chapter 9) (Stirzaker et al., 2015). In this study, there were three possible associations examined using epigenetics: differences between normal breast tissue and TNBC, differences between lymph node positive tumours and lymph node negative tumours and correlation between survival and epigenetics. Methylation differences were identified to predict prognosis in TNBC patients and to distinguish normal from tumour tissue, but no markers were identified to enable prediction of lymph node status from the primary tumour.

There may be a number of possible causes of the failure to predict lymph node status in breast cancer from characteristics of the primary tumour.

- Incorrect selection of biomarkers

More than one thousand biomarkers have been described that associate with lymph node status in breast cancer, however, none of these have been implemented due to failure to

validate in separate patient cohorts or generalise for sporadic breast cancers. There are also usually conflicting results for each biomarker with some studies showing an association with lymph node status and other studies showing a lack of association with lymph node status.

The large number of variables suggests that there are no single strong classifiers associated with lymph node status.

This project evaluated protein expression by IHC and this methodology is used in routine practice for clinical decisions. In breast carcinoma, IHC for oestrogen receptor and progesterone receptor forms part of the standard of care for management of breast cancer patients. The IHC assay replaced the radioligand binding assay which was originally used for assessment of oestrogen receptor status. The IHC assays have been extensively evaluated for clinical utility (Hammond et al., 2010) and the use of IHC would be expected to correlate with clinical characteristics of breast carcinomas and the response to treatment.

The lack of correlation between lymph node metastases and biomarkers also seems to apply to epigenetic markers and whilst miRNA markers have shown some association with lymph node status the results appear to be similar. A preliminary study did show some differences between miRNA expression in a primary breast compared to cells from lymphatic channels and lymph node metastases Appendix). Even using laser capture microdissection, the cells isolated from the lymphatic channels are not a uniform population, and if clusters are required for the development of metastases then these cells will represent an admixture of cell types with variable expression of biomarkers and a complex interaction of cell signalling.

The number of potential biomarkers makes it difficult to develop models as complete data covering such a large dataset is difficult to obtain even with such projects as TCGA. This means that selecting a relatively limited number of biomarkers is problematic as the there is a risk that the selected biomarkers will not associate with the clinical feature being evaluated. However, even with analysis of large datasets such as the TCGA with gene expression and epigenetic markers, a signature for lymph node metastasis has still not been identified.

In contrast, a number of different algorithms have been developed to predict survival and even though these use different input variables, the results appear to reasonably similar. This suggests that for survival there are multiple markers that correlate with survival and

that multiple algorithms are able to lead to the same conclusion. The same does not appear to hold true for lymph node status, as if the same process occurred, then it should be possible to develop models to predict lymph node status using a number of different methodologies, different biomarkers and algorithms which would result in a similar outcome.

- Inability of protein expression to correlate with clinical outcome

IHC is an antigen-antibody reaction used as a measurement of protein expression in tissue. However, IHC does not show a linear correlation with protein expression as the method of detection is not linear. With enhanced sensitivity the IHC reaction approaches a dichotomous reaction of positive and negative. IHC will therefore not correlate with subtle changes in protein expression and this has the potential to impact on associations with clinical characteristics such as lymph node status if there are only minor changes in the biomarkers that are associated with a particular feature. However, gene expression changes and methylation changes have also not shown a correlation with lymph node status and these methodologies have a more dynamic range.

- ANN may not be able to develop a successful model

ANNs are a particular type of model system that has been used in biological systems and other inherently unpredictable systems, such as share market predictions and image analysis, because of inherently noisy data. Most statistical analyses assume that the input data is accurate and correct, however, in biological systems there are multiple sources of variation in the data. For IHC, the data is subject to interpretation and there is less than 100% concordance between interpretation of IHC even when other variables are controlled. The process of tissue fixation and IHC introduces variation in the staining intensity of the biomarker and these variables are not uniform and are not recorded. An ANN was selected as the best model because no assumptions are made about the data accuracy. Input variables are used with a defined output variable and each input variable is evaluated to determine if it is useful in the models tested. Different levels of noise were selected including noisy, moderately noisy and extremely noisy, but no successful model was able to be developed for lymph node status. In contrast, a model was developed of

breast cancer specific survival which would suggest that selection of an ANN to develop a model for a biological system is a reasonable choice.

- Tumour heterogeneity

Whilst tumours are a clonal proliferation of neoplastic cells, all tumours are heterogeneous. Any methodology assessing tumour biomarkers is usually a composite of tumour cells and also a varying number of normal cells. IHC is assessed on a morphological section and the expression of the protein is correlated with the different cell compartments and is directly linked with tumour. Gene expression profiling and other molecular testing is often performed on fresh/frozen tissue. This tissue is used because of the preservation of nucleic acids, but it is difficult to assess the morphology of the tissue being tested. This variation in tumour content may potentially account for the lack of the ability to predict lymph node status using such techniques as gene expressing profiling and DNA methylation. Results are also confounded by the requirement for a normal control sample and the selection of a control may influence the results. Even in FFPE material where the amount of tumour content can be assessed, the population of cells is not homogeneous. These factors have the potential to dilute the expression of biomarkers that may be associated with lymph node metastases and may account for the inability of models to predict nodal status.
However, IHC expression is assessed in the tumour cells and scoring methods do take into account variation in the expression of the particular biomarker with an assessment of percentage of tumour cells staining and the intensity of staining.
Against this hypothesis is the fact that molecular classification of breast carcinomas has been successfully performed despite these inherent limitations and variations in samples, and similarly models have been developed for molecular grading and prediction of prognosis.

- Development of lymph node metastases may not be due to molecular or protein changes in the primary tumour

There are two models proposed for tumour progression, the linear progression model and the parallel progression model (Kimbung et al., 2015). Traditionally metastasis has been associated with the linear progression model which assumes a step-wise progression of cancer. This progression is due to the accumulation of sequential genetic abnormalities

and the eventual development of a metastatic phenotype. In support of this model is the association of tumour size with metastases and the similarity of gene expression signatures between primary tumours and the metastasis (Kimbung et al., 2015). In contrast, the parallel progression model proposes that metastatic potential in a tumour is acquired early in the development of a tumour and that breast carcinoma is a systemic disease from the outset (Kimbung et al., 2015). If the parallel progression model is the correct model for breast cancer progression, at least for lymph node metastases, then this would be a possible explanation for the inability to predict lymph node status from characteristics of the primary tumour. Even very small tumours would have expression patterns and phenotypic and genetic characteristics associated with lymph node metastases and these would not be acquired in a stepwise fashion.

This theory is also supported by the data from circulating tumour cells where clumps of tumour cells may be required to initiate and result in metastatic lesions and this may also apply to lymph node metastases. If this is the situation then the cells being analysed will be a mixture of cell types with different biological properties and any assessment of those cells will be a combined analysis rather than a pure cell population.

This is analogous to the molecular profiles seen in ductal carcinoma in situ (DCIS) and invasive ductal breast carcinoma. Molecular profiling of infiltrating ductal carcinoma has not shown a difference in the molecular profile compared to DCIS (Abba et al., 2015).

Statistical modelling has also been used to predict metastatic disease (Michor et al., 2006) and a stochastic nature of breast cancer progression accounts for some of the observed characteristics of breast cancer progression (Vanharanta and Massague, 2013).

Tumours undergo evolution with the acquisition of an increasing number of genetic abnormalities and some of these will result in a metastatic phenotype although not every metastatic phenotype will necessarily be reached by the same evolutionary route (Vanharanta and Massague, 2013). Even with the dissemination of tumour cells with metastatic potential, only a small number of cells actually form a metastasis. A stochastic model would explain the existing genetic evidence where tumour cells have the ability to form metastasis, but at a low frequency (Vanharanta and Massague, 2013).

Genetic analysis has been performed on multiple tumour types and this has also included an analysis of metastatic lesions. Differences in the mutational profile have been identified in metastases and primary tumours and between different metastatic sites, however metastasis-specific mutations have not been identified (Vanharanta and Massague, 2013).

This suggests that metastasis is not a distinct genetic event defined by a specific mutation or combination of easily definable mutations and whilst this is usually applied to distant metastases, the same issues appear to apply to lymph node metastases. The process of metastasis may be too complex to be defined a limited number of mutations and instead rely on a complex interaction between EMT of the tumour cells, epigenetic changes, tumour phenotype and the interaction between different tumour cells and the microenvironment.

The latter theory is probably the most likely explanation for the failure of the ANN to develop a model to predict lymph node metastasis in breast cancer. Different approaches were used to define appropriate biomarkers that could be used for the development of the ANN model. These variables included clinicopathological characteristics of the primary tumour such as tumour size. Even for tumour size, different approaches were considered such as maximum tumour size, a calculation of tumour volume using an ellipsoid model and volume using a rectangular tumour model. Protein expression for multiple markers was assessed using immunohistochemistry as well as miRNA (Wee et al., 2012) and epigenetic markers (Stirzaker et al., 2015).
An analysis of a breast cancer cohort also failed to show any differences in the tumours from the cluster compared to the tumours from case-matched controls (Waddell et al., 2012).
An ANN has been used to develop a model for lymph node metastases in oesophageal squamous cell carcinoma using gene expression profiling (Kan et al., 2004). In this study the ANN predicted lymph node metastasis in 10 of 13 validation cases (77%) and in 24 of 28 (86%) for the entire patient cohort with a sensitivity of 88% and a specificity of 82% (Kan et al., 2004). Interestingly the authors of this article thought that the ANN model was useful for the clinical prediction of lymph node status and could be applied clinically, however the sensitivity and specificity of the model does not appear to be sufficiently high enough for implementation.
Models for the prediction of lymph node metastases for different tumour types have not been developed with a high degree of accuracy. Standard histological assessment of lymphovascular invasion in tumours is associated with lymph node metastases in multiple tumour types, but this has not been supplanted by any new advanced methodologies and

none have been implemented into routine diagnostic practice to improve the current pathological assessment of tumour features.

Given the extensive data on biomarkers reported to be associated with lymph node metastases, if there was a strong link between a particular biomarker and nodal status, it is likely that it would have been identified. The fact that there are so many reported biomarkers, gene expression signatures, epigenetic associations and associations with miRNA suggests that there is no single or limited number of biomarkers that correlates strongly with lymph node status. This would also provide support for the theory of a complex process of lymph node metastasis that cannot be readily be defined with a large number of tumour cells having metastatic potential, but only a small number producing metastatic disease due to mechanistic and microenvironmental interactions.

In contrast, it is easier to predict outcome in breast cancer patients as is indicated by the availability of commercial assays for selection of lymph node negative patients into low and high risk patients for selection of treatment options. These assays essentially use markers of proliferation and similar predictions can be made with IHC for Ki-67. Proliferation in the form of a mitotic count is incorporated into the grading system for breast cancer with the generation of three grades. A model was developed for the determination of molecular grade to stratify morphological Grade 2 tumours into predicted Grade 1 and Grade 3 tumours with a correlation with breast cancer specific survival (Francis et al., 2012). Similarly, an ANN model was developed to enable the prediction of breast cancer specific survival for individual breast cancer patients and epigenetic profiling was used to stratify triple-negative breast cancer patients into different prognostic groups (Stirzaker et al., 2015).

In tumours, it is postulated that distant metastases can arise by either direct haematogenous dissemination or by lymph node metastases and then subsequent haematogenous spread from the nodal metastases. These pathways are different in different tumour types. In breast cancer, lymph node metastases are usually present when distant metastases occur but they may not necessarily occur sequentially, but may occur simultaneously. If the process of distant metastases was correlated with an intermediary step of lymph node metastases prior to subsequent haematogenous dissemination, then it

would be expected that prediction of patient survival, which is associated with distant metastases, would also be able to predict lymph node status. This does not appear to be the cases as is illustrated in this thesis. The process of lymph node metastasis and distant metastases appear to be independent processes. Distant metastases, in the form of an adverse patient outcome can be predicted by different models using different biomarkers. This suggests that the metastatic phenotype can be defined by different biomarkers that correlate with an adverse outcome.

In contrast, no model has been able to be developed either in the published literature or by this project that is able to predict positive lymph node status in breast cancer with a significant high degree of accuracy that would enable implementation into clinical practice. Sentinel lymph node biopsy is currently used to assess the lymph node status in breast cancer and this relies and direct sample of the lymph node. Examination of the node is by direct histological examination or by molecular detection of cytokeratins and this is then used to select patients for further evaluation by an axillary lymph node dissection. The process of dissemination of tumour cells to the lymph nodes and the formation of lymph node metastases would appear to be separate to that for distant metastases and the mechanisms involved appear to be independent. The input variables used for development of an ANN for prediction of breast cancer specific survival were also included as input variables for the development of models for prediction of lymph node status. If the progression of breast cancer followed a sequential process from lymph node positive disease to distant metastases it is likely that some of the markers associated with distant metastases would also be associated with lymph node status. An alternative hypothesis is that subsequent evolutionary changes occur in the lymph node metastasis that then result in distant metastases. Removal of the lymph nodes in breast cancer is used as a prognostic marker, but it also possible that removal of positive nodes reduces the subsequent development of distant metastases. However, when axillary lymph node dissection was routinely performed on all breast cancer patients, a proportion of patients with no lymph node metastases still went on to develop distant metastases, although at a lower frequency. Approximately one third of women with lymph-node negative breast cancer will develop distant metastases (Weigelt et al., 2005b). If lymph node metastases were required for tumour evolution prior to distant dissemination, then it would be expected that the incidence of distant metastases in lymph node negative patients would be

significantly lower. This again supports the concept that lymph node metastases and distant metastases are independent events.

Chapter 11

Conclusion

11.1. Conclusion

Lymph node status is an independent prognostic marker in breast cancer. Whilst the standard of care has changed from axillary lymph node dissection to sentinel lymph node dissection, there is still an associated morbidity associated with the latter procedure.
An artificial neural network was used to attempt to develop a model to predict lymph node status from characteristics and biomarkers from the primary tumour. This would result in a reduction of morbidity whist retaining the prognostic usefulness of lymph node status.
An ANN model was not able to be successfully developed to predict lymph node status, although models were developed to predict the molecular grade of breast cancer and to individually predict breast cancer specific survival.
The data from this project suggests that it is not possible to predict lymph node status form the primary tumour characteristics with sufficient clinical accuracy to forgo sentinel lymph node biopsy. It also suggests that lymph node metastases and distant metastases are separate independent events. Lymph node metastases are associated with a complex phenotype that is not defined a single genetic mutation and the elucidation of the events resulting in this biological cascade are still be to be defined.

Chapter 12

Bibliography

2011. miRBase - http://www.mirbase.org/.
AALTONEN, K., AHLIN, C., AMINI, R. M., SALONEN, L., FJALLSKOG, M. L., HEIKKILA, P., NEVANLINNA, H. & BLOMQVIST, C. 2006. Reliability of cyclin A assessment on tissue microarrays in breast cancer compared to conventional histological slides. *Br J Cancer*.
ABBA, M. C., GONG, T., LU, Y., LEE, J., ZHONG, Y., LACUNZA, E., BUTTI, M., TAKATA, Y., GADDIS, S., SHEN, J., ESTECIO, M. R., SAHIN, A. A. & ALDAZ, C. M. 2015. A Molecular Portrait of High-Grade Ductal Carcinoma In Situ. *Cancer Res,* 75**,** 3980-90.
ABDUL-RASOOL, S., KIDSON, S. H., PANIERI, E., DENT, D., PILLAY, K. & HANEKOM, G. S. 2006. An evaluation of molecular markers for improved detection of breast cancer metastases in sentinel nodes. *J Clin Pathol,* 59**,** 289-97.
ACETO, N., BARDIA, A., MIYAMOTO, D. T., DONALDSON, M. C., WITTNER, B. S., SPENCER, J. A., YU, M., PELY, A., ENGSTROM, A., ZHU, H., BRANNIGAN, B. W., KAPUR, R., STOTT, S. L., SHIODA, T., RAMASWAMY, S., TING, D. T., LIN, C. P., TONER, M., HABER, D. A. & MAHESWARAN, S. 2014. Circulating tumor cell clusters are oligoclonal precursors of breast cancer metastasis. *Cell,* 158**,** 1110-22.
ADEYINKA, A., MERTENS, F., BONDESON, L., GARNE, J. P., BORG, A., BALDETORP, B. & PANDIS, N. 2000. Cytogenetic heterogeneity and clonal evolution in synchronous bilateral breast carcinomas and their lymph node metastases from a male patient without any detectable BRCA2 germline mutation. *Cancer Genet Cytogenet,* 118**,** 42-7.
AGOFF, S. N., SWANSON, P. E., LINDEN, H., HAWES, S. E. & LAWTON, T. J. 2003. Androgen receptor expression in estrogen receptor-negative breast cancer. Immunohistochemical, clinical, and prognostic associations. *Am J Clin Pathol,* 120**,** 725-31.
AHLGREN, J., STAL, O., WESTMAN, G. & ARNESSON, L. G. 1994. Prediction of axillary lymph node metastases in a screened breast cancer population. South-East Sweden Breast Cancer Group. *Acta Oncol,* 33**,** 603-8.
AHR, A., HOLTRICH, U., SOLBACH, C., SCHARL, A., STREBHARDT, K., KARN, T. & KAUFMANN, M. 2001. Molecular classification of breast cancer patients by gene expression profiling. *J Pathol,* 195**,** 312-20.
AHR, A., KARN, T., SOLBACH, C., SEITER, T., STREBHARDT, K., HOLTRICH, U. & KAUFMANN, M. 2002. Identification of high risk breast-cancer patients by gene expression profiling. *Lancet,* 359**,** 131-2.
AIHW & NBCC 2006. Breast cancer in Australia: an overview, 2006. . *Cancer series no. 34. cat. no. CAN 29.* Canberra: AIHW.
AILLES, L. E. & WEISSMAN, I. L. 2007. Cancer stem cells in solid tumors. *Curr Opin Biotechnol,* 18**,** 460-6.
AL-HAJJ, M. & CLARKE, M. F. 2004. Self-renewal and solid tumor stem cells. *Oncogene,* 23**,** 7274-82.
AL-HAJJ, M., WICHA, M. S., BENITO-HERNANDEZ, A., MORRISON, S. J. & CLARKE, M. F. 2003. Prospective identification of tumorigenic breast cancer cells. *Proc Natl Acad Sci U S A,* 100**,** 3983-8.
AL KURAYA, K., SIMON, R. & SAUTER, G. 2004. Tissue microarrays for high-throughput molecular pathology. *Ann Saudi Med,* 24**,** 169-74.
ALITALO, K. 2011. The lymphatic vasculature in disease. *Nat Med,* 17**,** 1371-80.
ALLRED, D. C., WU, Y., MAO, S., NAGTEGAAL, I. D., LEE, S., PEROU, C. M., MOHSIN, S. K., O'CONNELL, P., TSIMELZON, A. & MEDINA, D. 2008. Ductal carcinoma in situ and the emergence of diversity during breast cancer evolution. *Clin Cancer Res,* 14**,** 370-8.
ALRAN, S., DE RYCKE, Y., FOURCHOTTE, V., CHARITANSKY, H., LAKI, F., FALCOU, M. C., BENAMOR, M., FRENEAUX, P., SALMON, R. J. & SIGAL-ZAFRANI, B. 2007. Validation and limitations of use of a breast cancer nomogram predicting the likelihood of non-sentinel node involvement after positive sentinel node biopsy. *Ann Surg Oncol,* 14**,** 2195-201.
ALVAGER, T., SMITH, T. J. & VIJAI, F. 1994. The Use of Artificial Neural networks in biomedical Technologies: An Introduction. *Biomedical Instrumentation & Technology,* 28**,** 315-322.
ANAN, K., MITSUYAMA, S., TAMAE, K., NISHIHARA, K., IWASHITA, T., ABE, Y., IHARA, T. & TOYOSHIMA, S. 2000. Axillary lymph node metastases in patients with small carcinomas of the breast: is accurate prediction possible? *Eur J Surg,* 166**,** 610-5.
ANNA, S., CATERINA, M., REBECCA, S., ISABELLA, C., LUIGIA, M., PAOLA, C., GIAMPIERO, G., MILENA, C., ENRICO, D. A. & GIANNI, B. 2006. Routine assessment of prognostic factors in breast cancer using a multicore tissue microarray procedure. *Virchows Archiv,* V449**,** 288-296.
ARNAOUT-ALKARAIN, A., KAHN, H. J., NAROD, S. A., SUN, P. A. & MARKS, A. N. 2007. Significance of lymph vessel invasion identified by the endothelial lymphatic marker D2-40 in node negative breast cancer. *Mod Pathol,* 20**,** 183-91.
ARNOULD, L., DENOUX, Y., MACGROGAN, G., PENAULT-LLORCA, F., FICHE, M., TREILLEUX, I., MATHIEU, M. C., VINCENT-SALOMON, A., VILAIN, M. O. & COUTURIER, J. 2003. Agreement

between chromogenic in situ hybridisation (CISH) and FISH in the determination of HER2 status in breast cancer. *British J of Cancer,* 88**,** 1587-1591.

ASHTURKAR, A. V., PATHAK, G. S., DESHMUKH, S. D. & PANDAVE, H. T. 2011. Factors predicting the axillary lymph node metastasis in breast cancer: is axillary node clearance indicated in every breast cancer patient?: factors predicting the axillary lymphnode metastases in breast cancer. *Indian J Surg,* 73**,** 331-5.

ASTION, M. L. & WILDING, P. 1992a. The Application of Backpropagation Neural Networks to Problems in Pathology and Laboratory Medicine. *Arch Pathol Lab Med,* 116**,** 995-1001.

ASTION, M. L. & WILDING, P. 1992b. Application of Neural Networks to the Interpretation of Laboratory Data in Cancer Diagnosis. *Clinical Chemistry,* 38**,** 34-38.

AUCHINCLOSS, H. 1963. Significance of Location and Number of Axillary Metastases in Carcinoma of the Breast. *Ann Surg,* 158**,** 37-46.

BACKUS, J., LAUGHLIN, T., WANG, Y., BELLY, R., WHITE, R., BADEN, J., JUSTUS MIN, C., MANNIE, A., TAFRA, L., ATKINS, D. & VERBANAC, K. M. 2005. Identification and characterization of optimal gene expression markers for detection of breast cancer metastasis. *J Mol Diagn,* 7**,** 327-36.

BAFFA, R., FASSAN, M., VOLINIA, S., O'HARA, B., LIU, C. G., PALAZZO, J. P., GARDIMAN, M., RUGGE, M., GOMELLA, L. G., CROCE, C. M. & ROSENBERG, A. 2009. MicroRNA expression profiling of human metastatic cancers identifies cancer gene targets. *J Pathol,* 219**,** 214-21.

BALIC, M., LIN, H., YOUNG, L., HAWES, D., GIULIANO, A., MCNAMARA, G., DATAR, R. H. & COTE, R. J. 2006. Most early disseminated cancer cells detected in bone marrow of breast cancer patients have a putative breast cancer stem cell phenotype. *Clin Cancer Res,* 12**,** 5615-21.

BARTLETT, J., GOING, J., MALLON, E., EWATTERS, A., REEVES, J., STANTON, P., RIVHMOND, J., DONALD, B., FERRIER, R. & COOKE, T. 2001. Evaluating HER2 amplification and overexpression in breast cancer. *Journal of Pathology,* 195**,** 422-428.

BASSI, K. K., SEENU, V., BALLEHANINNA, U. K., PARSHAD, R., CHUMBER, S., DHAR, A., GUPTA, S. D., KUMAR, R. & SRIVASTAVA, A. 2006. Second echelon node predicts metastatic involvement of additional axillary nodes following sentinel node biopsy in early breast cancer. *Indian J Cancer,* 43**,** 103-9.

BATHEN, T. F., JENSEN, L. R., SITTER, B., FJOSNE, H. E., HALGUNSET, J., AXELSON, D. E., GRIBBESTAD, I. S. & LUNDGREN, S. 2007. MR-determined metabolic phenotype of breast cancer in prediction of lymphatic spread, grade, and hormone status. *Breast Cancer Res Treat,* 104**,** 181-9.

BATTIFORA, H. 1986. The multitumor (sausage) tissue block: novel method for immunohistochemical antibody testing. *Lab Invest,* 55**,** 244-8.

BEDROSIAN, I., REYNOLDS, C., MICK, R., CALLANS, L. S., GRANT, C. S., DONOHUE, J. H., FARLEY, D. R., HELLER, R., CONANT, E., OREL, S. G., LAWTON, T., FRAKER, D. L. & CZERNIECKI, B. J. 2000. Accuracy of sentinel lymph node biopsy in patients with large primary breast tumors. *Cancer,* 88**,** 2540-5.

BERTUCCI, F., HOULGATTE, R., BENZIANE, A., GRANJEAUD, S., ADELAIDE, J., TAGETT, R., LORIOD, B., JACQUEMIER, J., VIENS, P., JORDAN, B., BIRNBAUM, D. & NGUYEN, C. 2000. Gene expression profiling of primary breast carcinomas using arrays of candidate genes. *Hum Mol Genet,* 9**,** 2981-91.

BEVILACQUA, J. L., KATTAN, M. W., FEY, J. V., CODY, H. S., 3RD, BORGEN, P. I. & VAN ZEE, K. J. 2007. Doctor, what are my chances of having a positive sentinel node? A validated nomogram for risk estimation. *J Clin Oncol,* 25**,** 3670-9.

BILOUS, M., MOREY, A., ARMES, J., CUMMINGS, M. & FRANCIS, G. 2006. Chromogenic in situ hybridisation testing for HER2 gene amplification in breast cancer produces highly reproducible results concordant with fluorescence in situ hybridisation and immunohistochemistry. *Pathology,* 38**,** 120-4.

BIRRELL, S. N., BENTEL, J. M., HICKEY, T. E., RICCIARDELLI, C., WEGER, M. A., HORSFALL, D. J. & TILLEY, W. D. 1995. Androgens induce divergent proliferative responses in human breast cancer cell lines. *J Steroid Biochem Mol Biol,* 52**,** 459-67.

BLOOM, H. J. & RICHARDSON, W. W. 1957. Histological grading and prognosis in breast cancer; a study of 1409 cases of which 359 have been followed for 15 years. *Br J Cancer,* 11**,** 359-77.

BOCKHORN, M., JAIN, R. K. & MUNN, L. L. 2007. Active versus passive mechanisms in metastasis: do cancer cells crawl into vessels, or are they pushed? *Lancet Oncol,* 8**,** 444-8.

BONASTRE, J., MARGUET, S., LUEZA, B., MICHIELS, S., DELALOGE, S. & SAGHATCHIAN, M. 2014. Cost effectiveness of molecular profiling for adjuvant decision making in patients with node-negative breast cancer. *J Clin Oncol,* 32**,** 3513-9.

BONO, P., WASENIUS, V. M., HEIKKILA, P., LUNDIN, J., JACKSON, D. G. & JOENSUU, H. 2004. High LYVE-1-positive lymphatic vessel numbers are associated with poor outcome in breast cancer. *Clin Cancer Res,* 10**,** 7144-9.

BORGSTEIN, P., MEIJER, S. & PIJPERS, R. 1997. Intradermal blue dye to identify sentinel lymphnode in breast cancer. *The Lancet,* 349**,** 1668-1669.

BOSTWICK, D. G. & BURKE, H. B. 2001. Prediction of individual patient outcome in cancer: Comparison of artificial neural networks and Kaplan-Meier methods. *Cancer,* 91**,** 1643-1646.

BOTTOS, A. & HYNES, N. E. 2014. Cancer: Staying together on the road to metastasis. *Nature,* 514**,** 309-10.

BOURDES, V. S., BONNEVAY, S., LISBOA, P. J. G., AUNG, M. S. H., CHABAUD, S., BACHELOT, T., PEROL, D. & NEGRIER, S. Breast Cancer Predictions by Neural Networks Analysis: a Comparison with Logistic Regression. Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE, 2007. 5424-5427.

BOWER, H. 1997. Identifying sentinel node could reduce surgery in breast cancer. *BMJ,* 315**,** 9.

BRAUN, M., FLUCKE, U., DEBALD, M., WALGENBACH-BRUENAGEL, G., WALGENBACH, K. J., HOLLER, T., POLCHER, M., WOLFGARTEN, M., SAUERWALD, A., KEYVER-PAIK, M., KUHR, M., BUTTNER, R. & KUHN, W. 2007. Detection of lymphovascular invasion in early breast cancer by D2-40 (podoplanin): a clinically useful predictor for axillary lymph node metastases. *Breast Cancer Res Treat*.

BRYAN, R. M., MERCER, R. J., BENNETT, R. C., RENNIE, G. C., LIE, T. H. & MORGAN, F. J. 1984. Androgen receptors in breast cancer. *Cancer,* 54**,** 2436-40.

BRYNE, J. C., VALEN, E., TANG, M. H., MARSTRAND, T., WINTHER, O., DA PIEDADE, I., KROGH, A., LENHARD, B. & SANDELIN, A. 2008. JASPAR, the open access database of transcription factor-binding profiles: new content and tools in the 2008 update. . *Nucleic Acids Res***,** D102-106.

BURKE, H. B. 1994. Artificial Neural Networks for Cancer Research: Outcome Prediction. *Seminars in Surgical Oncology,* 10**,** 73-79.

BURKE, H. B., GOODMAN, P. H., ROSEN, D. B., HANSON, D. E., WELNSTEIN, J., HARRELL, F., MARKS, J., WINCHESTER, D. & BOSTWICK, D. G. 1997. Artifical neural networks improve the accuracy of cancer survival prediction. *Cancer,* 79**,** 857-862.

BURKE, H. B. & HENSON, D. E. 1997. Histologic grade as a prognostic factor in breast carcinoma. *Cancer,* 80**,** 1703-5; discussion 1706-7.

CALLAGY, G., CATTANEO, E., DAIGO, Y., HAPPERFIELD, L., BOBROW, L. G., PHAROAH, P. D. P. & CALDAS, C. 2003. Molecular classification of breast carcinomas using tissue microarrays. *Diagnostic Molecular Pathology,* 12**,** 27-34.

CAMP, R. L., CHARETTE, L. A. & RIMM, D. L. 2000. Validation of tissue microarray technology in breast carcinoma. *Lab Invest,* 80**,** 1943-9.

CAMPBELL, L. L. & POLYAK, K. 2007. Breast tumor heterogeneity: cancer stem cells or clonal evolution? *Cell Cycle,* 6**,** 2332-8.

CARIATI, M. & PURUSHOTHAM, A. D. 2008. Stem cells and breast cancer. *Histopathology,* 52**,** 99-107.

CARLSSON, J., NORDGREN, H., SJOSTROM, J., WESTER, K., VILLMAN, K., BENGTSSON, N. O., OSTENSTAD, B., LUNDQVIST, H. & BLOMQVIST, C. 2004. HER2 expression in breast cancer primary tumours and corresponding metastases. Original data and literature review. *Br J Cancer,* 90**,** 2344-8.

CARMON, M., OLSHA, O., SCHECTER, W. P., RAVEH, D., REINUS, C., HERSHKO, D. D., BELLER, U. & GOLOMB, E. 2006. The "Sentinel Chain": a new concept for prediction of axillary node status in breast cancer patients. *Breast Cancer Res Treat,* 97**,** 323-8.

CARMONA, F. J., VILLANUEVA, A., VIDAL, A., MUNOZ, C., PUERTAS, S., PENIN, R. M., GOMA, M., LUJAMBIO, A., PIULATS, J. M., MESIA, R., SANCHEZ-CESPEDES, M., MANOS, M., CONDOM, E., ECCLES, S. A. & ESTELLER, M. 2012. Epigenetic disruption of cadherin-11 in human cancer metastasis. *J Pathol,* 228**,** 230-40.

CASTELLANO, I., ALLIA, E., ACCORTANZO, V., VANDONE, A. M., CHIUSA, L., ARISIO, R., DURANDO, A., DONADIO, M., BUSSOLATI, G., COATES, A. S., VIALE, G. & SAPINO, A. 2010. Androgen receptor expression is a significant prognostic factor in estrogen receptor positive breast cancers. *Breast Cancer Res Treat*.

CAVALLI, L. R. 2009. Molecular markers of breast axillary lymph node metastasis. *Expert Rev Mol Diagn,* 9**,** 441-54.

CAVALLI, L. R., CAVALIERI, L. M., RIBEIRO, L. A., CAVALLI, I. J., SILVEIRA, R. & ROGATTO, S. R. 1997. Cytogenetic evaluation of 20 primary breast carcinomas. *Hereditas,* 126**,** 261-8.

CAVALLI, L. R., MAN, Y. G., SCHWARTZ, A. M., RONE, J. D., ZHANG, Y., URBAN, C. A., LIMA, R. S., HADDAD, B. R. & BERG, P. E. 2008. Amplification of the BP1 homeobox gene in breast cancer. *Cancer Genet Cytogenet,* 187**,** 19-24.

CAVALLI, L. R., URBAN, C. A., DAI, D., DE ASSIS, S., TAVARES, D. C., RONE, J. D., BLEGGI-TORRES, L. F., LIMA, R. S., CAVALLI, I. J., ISSA, J. P. & HADDAD, B. R. 2003. Genetic and epigenetic alterations in sentinel lymph nodes metastatic lesions compared to their corresponding primary breast tumors. *Cancer Genet Cytogenet,* 146**,** 33-40.

CHAGPAR, A. B., SCOGGINS, C. R., MARTIN, R. C., 2ND, CARLSON, D. J., LAIDLEY, A. L., EL-EID, S. E., MCGLOTHIN, T. Q. & MCMASTERS, K. M. 2006. Prediction of sentinel lymph node-only disease in women with invasive breast cancer. *Am J Surg,* 192**,** 882-7.

CHANG, E., LEE, A., LEE, E., LEE, H., SHIN, O., OH, S. & KANG, C. 2004. HER-2/neu oncogene amplification by chromogenic in situ hybridization in 130 breast cancers using tissue microarray and clinical follow-up studies. *J Korean Med Sci,* 19**,** 390-6.

CHANG, J. C. 2007. HER2 Inhibition: From Discovery to Clinical Practice. *Clin Cancer Res,* 13**,** 1-3.

CHEN, S., SUPAKAR, P. C., VELLANOWETH, R. L., SONG, C. S., CHATTERJEE, B. & ROY, A. K. 1997. Functional role of a conformationally flexible homopurine/homopyrimidine domain of the androgen receptor gene promoter interacting with Sp1 and a pyrimidine single strand DNA-binding protein. *Mol Endocrinol,* 11**,** 3-15.

CHEN, S. L., HOEHNE, F. M. & GIULIANO, A. E. 2007. The prognostic significance of micrometastases in breast cancer: a SEER population-based analysis. *Ann Surg Oncol,* 14**,** 3378-84.

CHIN, S. F., WANG, Y., THORNE, N. P., TESCHENDORFF, A. E., PINDER, S. E., VIAS, M., NADERI, A., ROBERTS, I., BARBOSA-MORAIS, N. L., GARCIA, M. J., IYER, N. G., KRANJAC, T., ROBERTSON, J. F., APARICIO, S., TAVARE, S., ELLIS, I., BRENTON, J. D. & CALDAS, C. 2007. Using array-comparative genomic hybridization to define molecular portraits of primary breast cancers. *Oncogene,* 26**,** 1959-70.

CHO, E. Y., HAN, J. J., CHOI, Y. L., KIM, K. M. & OH, Y. L. 2008. Comparison of Her-2, EGFR and cyclin D1 in primary breast cancer and paired metastatic lymph nodes: an immunohistochemical and chromogenic in situ hybridization study. *J Korean Med Sci,* 23**,** 1053-61.

CHOI, W. W., LEWIS, M. M., LAWSON, D., YIN-GOEN, Q., BIRDSONG, G. G., COTSONIS, G. A., COHEN, C. & YOUNG, A. N. 2005. Angiogenic and lymphangiogenic microvessel density in breast carcinoma: correlation with clinicopathologic parameters and VEGF-family gene expression. *Mod Pathol,* 18**,** 143-52.

CHU, J. H., SUN, Z. Y., MENG, X. L., WU, J. H., HE, G. L., LIU, G. M. & JIANG, X. R. 2006. Differential metastasis-associated gene analysis of prostate carcinoma cells derived from primary tumor and spontaneous lymphatic metastasis in nude mice with orthotopic implantation of PC-3M cells. *Cancer Lett,* 233**,** 79-88.

CICCHETTI, D. V. 1992. Neural Networks and Diagnosis in the Clinical Laboratory: State of the Art. *Clinical Chemistry,* 38**,** 9-10.

CLIMENT, J., GARCIA, J. L., MAO, J. H., ARSUAGA, J. & PEREZ-LOSADA, J. 2007. Characterization of breast cancer by array comparative genomic hybridization. *Biochem Cell Biol,* 85**,** 497-508.

COBALEDA, C., CRUZ, J. J., GONZALEZ-SARMIENTO, R., SANCHEZ-GARCIA, I. & PEREZ-LOSADA, J. 2008. The Emerging Picture of Human Breast Cancer as a Stem Cell-based Disease. *Stem Cell Rev*.

CODY, H. S., 3RD 1999. Sentinel lymph node mapping in breast cancer. *Breast Cancer,* 6**,** 13-22.

COHEN, L. F., BRESLIN, T. M., KREURER, H. M., ROSS, M. I., HUNT, K. K. & SAHIN, A. A. 2000. Identification and evaluation of axillary sentinel lymph nodes in patients with breast carcinoma treated with neoadjuvant chemotheraphy. *The American Journal of Surgery,* 24**,** 1266-1272.

COHNHEIM, V. 1875. Congenitales, quergestreiftes Muskelsarkom der Nieren. *Virchows Archiv fur Pathologische Anatomie und Physiologie und fur Klinische Medizin,* 65**,** 64-9.

COOLEN, M. W., STATHAM, A. L., GARDINER-GARDEN, M. & CLARK, S. J. 2007. Genomic profiling of CpG methylation and allelic specificity using quantitative high-throughput mass spectrometry: critical evaluation and improvements. *Nucleic Acids Res,* 35**,** e119.

COUTANT, C., UZAN, S., MOREL, O., MALARTIC, C., AKERMAN, G. & BARRANGER, E. 2008. An axillary score (Barranger Score) to predict nonsentinel node metastasis in breast cancer patients with positive sentinel node. *Am J Surg,* 195**,** 135-6.

COX, C. E., SOLANGE, PENDAS, COX, J. M., JOSEPH, E., SHONS, A. R., YEATMAN, T., KU, N. N., LYMAN, G. H., BERMAN, C., HADDAD, F. & REINTGEN, D. S. 1998. Guidelines for Sentinel Node Biopsy and Lymphatic Mapping of Patients with Breast Cancer. *Annals of Surgery,* 227**,** 645-653.

COX, D. G., BLANCHE, H., PEARCE, C. L., CALLE, E. E., COLDITZ, G. A., PIKE, M. C., ALBANES, D., ALLEN, N. E., AMIANO, P., BERGLUND, G., BOEING, H., BURING, J., BURTT, N., CANZIAN, F.,

CHANOCK, S., CLAVEL-CHAPELON, F., FEIGELSON, H. S., FREEDMAN, M., HAIMAN, C. A., HANKINSON, S. E., HENDERSON, B. E., HOOVER, R., HUNTER, D. J., KAAKS, R., KOLONEL, L., KRAFT, P., LEMARCHAND, L., LUND, E., PALLI, D., PEETERS, P. H., RIBOLI, E., STRAM, D. O., THUN, M., TJONNELAND, A., TRICHOPOULOS, D. & YEAGER, M. 2006. A comprehensive analysis of the androgen receptor gene and risk of breast cancer: results from the National Cancer Institute Breast and Prostate Cancer Cohort Consortium (BPC3). *Breast Cancer Res,* 8**,** R54.

CROCITTO, L. E., HENDERSON, B. E. & COETZEE, G. A. 1997. Identification of two germline point mutations in the 5'UTR of the androgen receptor gene in men with prostate cancer. *J Urol,* 158**,** 1599-601.

CROSS, S. S., HARRISON, R. F. & KENNEDY, R. L. 1995. Introduction to neural networks. *The Lancet,* 346**,** 1075-1079.

CSERNI, G. 2007. Comparison of different validation studies on the use of the Memorial-Sloan Kettering Cancer Center nomogram predicting nonsentinel node involvement in sentinel node-positive breast cancer patients. *Am J Surg,* 194**,** 699-700.

CSERNI, G., BIANCHI, S., VEZZOSI, V., ARISIO, R., PETERSE, J. L., SAPINO, A., CASTELLANO, I., DRIJKONINGEN, M., KULKA, J., EUSEBI, V., FOSCHINI, M. P., BELLOCQ, J. P., MARIN, C., THORSTENSON, S., AMENDOEIRA, I., REINER-CONCIN, A., DECKER, T., LACERDA, M. & FIGUEIREDO, P. 2007. Validation of clinical prediction rules for a low probability of nonsentinel and extensive lymph node involvement in breast cancer patients. *Am J Surg,* 194**,** 288-93.

DALTON, L. W., PAGE, D. L. & DUPONT, W. D. 1994. Histologic grading of breast carcinoma. A reproducibility study. *Cancer,* 73**,** 2765-70.

DALTON, L. W., PINDER, S. E., ELSTON, C. E., ELLIS, I. O., PAGE, D. L., DUPONT, W. D. & BLAMEY, R. W. 2000. Histologic grading of breast cancer: linkage of patient outcome with level of pathologist agreement. *Mod Pathol,* 13**,** 730-5.

DANDACHI, N., DIETZE, O. & HAUSER-KRONBERGER, C. 2002. Chromogenic in situ hybridization: A novel approach to a practical and sensitive method for the detection of HER2 oncogene in archival human breast carcinoma. *Laboratory Investigation,* 82**,** 1007-1014.

DAUPHINE, C. E., HAUKOOS, J. S., VARGAS, M. P., ISAAC, N. M., KHALKHALI, I. & VARGAS, H. I. 2007. Evaluation of three scoring systems predicting non sentinel node metastasis in breast cancer patients with a positive sentinel node biopsy. *Ann Surg Oncol,* 14**,** 1014-9.

DE LAURENTIIS, M. & RAVDIN, P. M. 1994. Survival analysis of censored data: neural network analysis detection of complex interactions between variables. *Breast Cancer Res Treat,* 32**,** 113-8.

DEGNIM, A. C., REYNOLDS, C., PANTVAIDYA, G., ZAKARIA, S., HOSKIN, T., BARNES, S., ROBERTS, M. V., LUCAS, P. C., OH, K., KOKER, M., SABEL, M. S. & NEWMAN, L. A. 2005. Nonsentinel node metastasis in breast cancer patients: assessment of an existing and a new predictive nomogram. *Am J Surg,* 190**,** 543-50.

DETRE, S., JOTTI, G. S. & DOWSETL, M. 1995a.
A "quickscore" method for immunohistochemical semiquantitation: validation for oestrogen receptor in breast carcinomas. *J Clin Pathol,* 48**,** 876-878.

DETRE, S., SACLANI JOTTI, G. & DOWSETT, M. 1995b. A "quickscore" method for immunohistochemical semiquantitation: validation for oestrogen receptor in breast carcinomas. *J Clin Pathol,* 48**,** 876-8.

DHESY-THIND, B., PRITCHARD, K. I., MESSERSMITH, H., O'MALLEY, F., ELAVATHIL, L. & TRUDEAU, M. 2007. HER2/neu in systemic therapy for women with breast cancer: a systematic review. *Breast Cancer Res Treat*.

DIAZ, N. M. 2001. Laboratory Testing for HER2/neu in Breast Carcinoma: An Evolving Strategy to Predict Response to Targeted Therapy. *Cancer Control,* 8**,** pp 415 - 418.

DIETEL, M., ELLIS, I. O., HOFLER, H., KREIPE, H., MOCH, H., DANKOF, A., KOLBLE, K. & KRISTIANSEN, G. 2007. Comparison of automated silver enhanced in situ hybridisation (SISH) and fluorescence ISH (FISH) for the validation of HER2 gene status in breast carcinoma according to the guidelines of the American Society of Clinical Oncology and the College of American Pathologists. *Virchows Arch,* 451**,** 19-25.

DIXON, M. 1998. Sentinel node biopsy in breast cancer. *BMJ,* 317**,** 295-6.

DOANE, A. S., DANSO, M., LAL, P., DONATON, M., ZHANG, L., HUDIS, C. & GERALD, W. L. 2006. An estrogen receptor-negative breast cancer subset characterized by a hormonally regulated transcriptional program and response to androgen. *Oncogene,* 25**,** 3994-4008.

DONOHOE, E. J. 2001. Sentinel node imaging and biopsy in breast cancer patients. *The American Journal of Surgery,* 182**,** 426-428.

DOWLATSHAHI, K., FAN, M., SNIDER, H. C. & HABIB, F. A. 1997a. Lymph Node Micrometastases from Breast Carcinoma. *American Cancer Society,* 80**,** 1188-97.

DOWLATSHAHI, K., FAN, M., SNIDER, H. C. & HABIB, F. A. 1997b. Lymph node micrometastases from breast carcinoma: reviewing the dilemma. *Cancer,* 80**,** 1188-97.

ECCLES, S., PAON, L. & SLEEMAN, J. 2007. Lymphatic metastasis in breast cancer: importance and new insights into cellular and molecular mechanisms. *Clinical and Experimental Metastasis,* 24**,** 619-636.

ECCLES, S. A. & WELCH, D. R. 2007. Metastasis: recent discoveries and novel treatment strategies. *Lancet,* 369**,** 1742-57.

EGMONT-PETERSON, M., TALMON, J. L., BRENDER, J. & MCNAIR, P. 1994. On the quality of neural net classifers. *Artifical Intelligence in Medicine,* 6**,** 359-381.

EGUILUZ, C., VIGUERA, E., MILLAN, L. & PEREZ, J. 2006. Multitissue array review: A chronological description of tissue array techniques, applications and procedures. *Pathology - Research and Practice,* 202**,** 561-568.

EHRLICH, M. 2002. DNA methylation in cancer: too much, but also too little. *Oncogene,* 21**,** 5400-13.

ELEUTERI, A., TAGLIAFERRI, R., MILANO, L., DE PLACIDO, S. & DE LAURENTIIS, M. 2003. A novel neural network-based survival analysis model. *Neural Netw,* 16**,** 855-64.

ELKIN, E. B. & SCHNITT, S. J. 2004. In Reply. *J Clin Oncol,* 22**,** 4232-a-4233.

ELLIS, I., SCHNITT, S., SASTRE-GARAU, X., BUSSOLATI, G., TAVASSOLI, F., EUSEBI, V., PETERSE, J., MUKAI, K., TABAR, L., JACQUEMIER, J., CORNELISSE, C., SASCO, A., KAAKS, R., PISANI, P., GOLDGAR, D., DEVILEE, P., CLETON-JANSEN, M., BORRESEN-DALE, A., VEER, L. V. T. & SAPINO, A. 2003. Invasive breast carcinoma. *In:* TAVASSOLI, F. & DEVILEE, P. (eds.) *World Health Organization Classification of Tumours, Pathology and Genetics: Tumours of the Breast and Female Genital Organs.* Lyon: IARC Press.

ELLIS, I. O., GALEA, M., BROUGHTON, N., LOCKER, A., BLAMEY, R. W. & ELSTON, C. W. 1992. Pathological prognostic factors in breast cancer. II. Histological type. Relationship with survival in a large study with long-term follow-up. *Histopathology,* 20**,** 479-89.

ELLSWORTH, R. E., FIELD, L. A., LOVE, B., KANE, J. L., HOOKE, J. A. & SHRIVER, C. D. 2011. Differential gene expression in primary breast tumors associated with lymph node metastasis. *Int J Breast Cancer,* 2011**,** 142763.

ELSTON, C. W. & ELLIS, I. O. 1991. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. *Histopathology,* 19**,** 403-10.

EMERSON, J. C., SALMON, S. E., DALTON, W., MCGEE, D. L., YANG, J. M., THOMPSON, F. H. & TRENT, J. M. 1993. Cytogenetics and clinical correlations in breast cancer. *Adv Exp Med Biol,* 330**,** 107-18.

FABER, P. W., VAN ROOIJ, H. C., SCHIPPER, H. J., BRINKMANN, A. O. & TRAPMAN, J. 1993. Two different, overlapping pathways of transcription initiation are active on the TATA-less human androgen receptor promoter. The role of Sp1. *J Biol Chem,* 268**,** 9296-301.

FABER, P. W., VAN ROOIJ, H. C., VAN DER KORPUT, H. A., BAARENDS, W. M., BRINKMANN, A. O., GROOTEGOED, J. A. & TRAPMAN, J. 1991. Characterization of the human androgen receptor transcription unit. *J Biol Chem,* 266**,** 10743-9.

FENG, Y., SUN, B., LI, X., ZHANG, L., NIU, Y., XIAO, C., NING, L., FANG, Z., WANG, Y., ZHANG, L., CHENG, J., ZHANG, W. & HAO, X. 2006. Differentially expressed genes between primary cancer and paired lymph node metastases predict clinical outcome of node-positive breast cancer patients. *Breast Cancer Res Treat*.

FERRETTI, G., FELICI, A., PAPALDO, P., FABI, A. & COGNETTI, F. 2007. HER2/neu role in breast cancer: from a prognostic foe to a predictive friend. *Curr Opin Obstet Gynecol,* 19**,** 56-62.

FIDLER, I. J. 1991. The biology of cancer metastasis or, 'you cannot fix it if you do not know how it works'. *Bioessays,* 13**,** 551-4.

FILHO, O. M., IGNATIADIS, M. & SOTIRIOU, C. 2010. Genomic Grade Index: An important tool for assessing breast cancer tumor grade and prognosis. *Crit Rev Oncol Hematol*.

FILLMORE, C. M. & KUPERWASSER, C. 2008. Human breast cancer cell lines contain stem-like cells that self-renew, give rise to phenotypically diverse progeny and survive chemotherapy. *Breast Cancer Res,* 10**,** R25.

FISHER, B., JEONG, J. H., BRYANT, J., ANDERSON, S., DIGNAM, J., FISHER, E. R. & WOLMARK, N. 2004. Treatment of lymph-node-negative, oestrogen-receptor-positive breast cancer: long-term findings from National Surgical Adjuvant Breast and Bowel Project randomised clinical trials. *Lancet,* 364**,** 858-68.

FLOYD, C. E., LO, J. Y., YUN, A. J., SULLIVAN, D. C. & KORNGUTH, P. J. 1994. Prediction of Breast Cancer Malignancy Using an Artificial Neural Network. *Cancer,* 74**,** 2944-8.

FOSTER, R. S., JR. 1996. The biologic and clinical significance of lymphatic metastases in breast cancer. *Surg Oncol Clin N Am,* 5**,** 79-104.

FRANCIS, G. D., DIMECH, M., GILES, L. & HOPKINS, A. 2007a. Frequency and reliability of oestrogen receptor, progesterone receptor and HER2 in breast carcinoma determined by immunohistochemistry in Australasia: results of the RCPA Quality Assurance Program. *J Clin Pathol,* 60**,** 1277-83.

FRANCIS, G. D., DIMECH, M., GILES, L. & HOPKINS, A. 2007b. Frequency and reliability of oestrogen receptor, progesterone receptor and HER2 in breast carcinoma determined by immunohistochemistry in Australasia: results of the RCPA Quality Assurance Program. *J Clin Pathol,* 60**,** 1277-1283.

FRANCIS, G. D., JONES, M. A., BEADLE, G. F. & STEIN, S. R. 2009a. Bright-field in situ hybridization for HER2 gene amplification in breast cancer using tissue microarrays: correlation between chromogenic (CISH) and automated silver-enhanced (SISH) methods with patient outcome. *Diagn Mol Pathol,* 18**,** 88-95.

FRANCIS, G. D., JONES, M. A., BEADLE, G. F. & STEIN, S. R. 2009b. Bright-field in situ hybridization for HER2 gene amplification in breast cancer using tissue microarrays: correlation between chromogenic (CISH) and automated silver-enhanced (SISH) methods with patient outcome. *Diagn Mol Pathol. ,* 18**,** 88-95.

FRANCIS, G. D., STEIN, S. R. & FRANCIS, G. D. Prediction of histologic grade in breast cancer using an artificial neural network. Neural Networks (IJCNN), The 2012 International Joint Conference on, 10-15 June 2012 2012. 1-5.

FRIEDMAN, N. S. & FREEDMAN, M. D. 1994. Correlation of DNA flow cytometry and hormone receptors with axillary lymph node status in patients with carcinoma of the breast. *Md Med J,* 43**,** 963-5.

FRIEDRICH, K., WEBER, T., SCHEITHAUER, J., MEYER, W., HAROSKE, G., KUNZE, K. D. & BARETTON, G. 2008. Chromosomal genotype in breast cancer progression: comparison of primary and secondary manifestations. *Cell Oncol,* 30**,** 39-50.

FRIERSON, H. F., JR., WOLBER, R. A., BEREAN, K. W., FRANQUEMONT, D. W., GAFFEY, M. J., BOYD, J. C. & WILBUR, D. C. 1995. Interobserver reproducibility of the Nottingham modification of the Bloom and Richardson histologic grading scheme for infiltrating ductal carcinoma. *Am J Clin Pathol,* 103**,** 195-8.

FRKOVIC-GRAZIO, S. & BRACKO, M. 2002. Long term prognostic value of Nottingham histological grade and its components in early (pT1NOMO) breast carcinoma. *J Clin Pathol,* 55**,** 88-92.

FUJITA, H., KATAFUCHI, T., UEHARA, T. & NISHIMURA, T. 1992. Application of Artificial Neural Network to Computer-Aided Diagnosis of Coronary Artey Disease in Myocardial SPECT Bull's-eye Images. *The Journal of Nuclear Medicine,* 33**,** 272-276.

GANCBERG, D., JARVINEN, T., DI LEO, A., ROUAS, G., CARDOSO, F., PAESMANS, M., VERHEST, A., PICCART, M. J., ISOLA, J. & LARSIMONT, D. 2002a. Evaluation of HER-2/NEU protein expression in breast cancer by immunohistochemistry: an interlaboratory study assessing the reproducibility of HER-2/NEU testing. *Breast Cancer Research and Treatment,* 74**,** 113-120.

GANCBERG, D., LEO, A. D., ROUAS, G., JARVINEN, T., VERHEST, A., ISOLA, J., PICCART, M. J. & LARSIMONT, D. 2002b. Reliability of the tissue microarray based FISH for evaluationof the HER-2 oncogene in breast carcinoma. *J Clin Pathol,* 55.

GANN, P. H., COLILLA, S. A., GAPSTUR, S. M., WINCHESTER, D. J. & WINCHESTER, D. P. 1999. Factors associated with axillary lymph node metastasis from breast carcinoma: descriptive and predictive analyses. *Cancer,* 86**,** 1511-9.

GERVASONI, J. E., JR., TANEJA, C., CHUNG, M. A. & CADY, B. 2000. Biologic and clinical significance of lymphadenectomy. *Surg Clin North Am,* 80**,** 1631-73.

GILTNANE, J. M. & RIMM, D. L. 2004. Technology insight: Identification of biomarkers with tissue microarray technology. *Nat Clin Pract Oncol,* 1**,** 104-11.

GONZALEZ-ANGULO, A. M., STEMKE-HALE, K., PALLA, S. L., CAREY, M., AGARWAL, R., MERIC-BERSTAM, F., TRAINA, T. A., HUDIS, C., HORTOBAGYI, G. N., GERALD, W. L., MILLS, G. B. & HENNESSY, B. T. 2009. Androgen receptor levels and association with PIK3CA mutations and prognosis in breast cancer. *Clin Cancer Res,* 15**,** 2472-8.

GOTTE, M. 2010. MicroRNAs in breast cancer pathogenesis. *Minerva Ginecol,* 62**,** 559-71.

GRABAU, D., JENSEN, M. B., RANK, F. & BLICHERT-TOFT, M. 2007. Axillary lymph node micrometastases in invasive breast cancer: national figures on incidence and overall survival. *Apmis,* 115**,** 828-37.

GREY, S. R., DLAY, S. S., LEONE, B. E., CAJONE, F. & SHERBET, G. V. 2003. Prediction of nodal spread of breast cancer by using artificial neural network-based analyses of S100A4, nm23 and steroid receptor expression. *Clin Exp Metastasis,* 20**,** 507-14.

GRIFFITHS-JONES, S. miRBase: microRNA sequences and annotation. *Curr Protoc Bioinformatics,* Chapter 12**,** Unit 12 9 1-10.

GRIFFITHS-JONES, S. 2006. miRBase: the microRNA sequence database. *Methods Mol Biol,* 342**,** 129-38.
GRIFFITHS-JONES, S., GROCOCK, R. J., VAN DONGEN, S., BATEMAN, A. & ENRIGHT, A. J. 2006. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res,* 34**,** D140-4.
GRIFFITHS-JONES, S., SAINI, H. K., VAN DONGEN, S. & ENRIGHT, A. J. 2008. miRBase: tools for microRNA genomics. *Nucleic Acids Res,* 36**,** D154-8.
GUARNIERI, A., NERI, A., CORREALE, P. P., LOTTINI, M., TESTA, M., MARIANI, F., TUCCI, E., MEGHA, T., CINTORINO, M. & CARLI, A. 2001. Prediction of lymph node status by analysis of prognostic factors and possible indications for elective axillary dissection in T1 breast cancers. *Eur J Surg,* 167**,** 255-9.
GUNNINGHAM, S. P., CURRIE, M. J., HAN, C., ROBINSON, B. A., SCOTT, P. A., HARRIS, A. L. & FOX, S. B. 2001. VEGF-B expression in human primary breast cancers is associated with lymph node metastasis but not angiogenesis. *J Pathol,* 193**,** 325-32.
GUPTA, D., MIDDLETON, L. P., WHITAKER, M. J. & ABRAMS, J. 2003. Comparison of fluorescence and chromogenic in situ hybridization for detection of HER-2/neu oncogene in breast cancer. *Am J Clin Pathol,* 119**,** 381-7.
GUSTERSON, B. A., GELBER, R. D. & GOLDHIRSCH, A. 1992. Prognostic importance of c-erbB-2 expression in breast cancer. *J Clin Oncol,* 10.
GYORFFY, B., HATZIS, C., SANFT, T., HOFSTATTER, E., AKTAS, B. & PUSZTAI, L. 2015. Multigene prognostic tests in breast cancer: past, present, future. *Breast Cancer Res,* 17**,** 11.
HABEL, L. A., SAKODA, L. C., ACHACOSO, N., MA, X. J., ERLANDER, M. G., SGROI, D. C., FEHRENBACHER, L., GREENBERG, D. & QUESENBERRY, C. P., JR. 2013. HOXB13:IL17BR and molecular grade index and risk of breast cancer death among patients with lymph node-negative invasive disease. *Breast Cancer Res,* 15**,** R24.
HAMMERSTON, D. 1993. Neural networks at work. *IEEE Spectrum,* 30**,** 26-32.
HAMMOND, M. E., HAYES, D. F., DOWSETT, M., ALLRED, D. C., HAGERTY, K. L., BADVE, S., FITZGIBBONS, P. L., FRANCIS, G., GOLDSTEIN, N. S., HAYES, M., HICKS, D. G., LESTER, S., LOVE, R., MANGU, P. B., MCSHANE, L., MILLER, K., OSBORNE, C. K., PAIK, S., PERLMUTTER, J., RHODES, A., SASANO, H., SCHWARTZ, J. N., SWEEP, F. C., TAUBE, S., TORLAKOVIC, E. E., VALENSTEIN, P., VIALE, G., VISSCHER, D., WHEELER, T., WILLIAMS, R. B., WITTLIFF, J. L. & WOLFF, A. C. 2010. American Society of Clinical Oncology/College Of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer. *J Clin Oncol,* 28**,** 2784-95.
HAN, W., HAN, M. R., KANG, J. J., BAE, J. Y., LEE, J. H., BAE, Y. J., LEE, J. E., SHIN, H. J., HWANG, K. T., HWANG, S. E., KIM, S. W. & NOH, D. Y. 2006. Genomic alterations identified by array comparative genomic hybridization as prognostic markers in tamoxifen-treated estrogen receptor-positive breast cancer. *BMC Cancer,* 6**,** 92.
HANAHAN, D. & WEINBERG, R. A. 2000. The hallmarks of cancer. *Cell,* 100**,** 57-70.
HANAHAN, D. & WEINBERG, R. A. 2011. Hallmarks of cancer: the next generation. *Cell,* 144**,** 646-74.
HAO, X., SUN, B., HU, L., LAHDESMAKI, H., DUNMIRE, V., FENG, Y., ZHANG, S. W., WANG, H., WU, C., WANG, H., FULLER, G. N., SYMMANS, W. F., SHMULEVICH, I. & ZHANG, W. 2004. Differential gene and protein expression in primary breast malignancies and their lymph node metastases as revealed by combined cDNA microarray and tissue microarray analysis. *Cancer,* 100**,** 1110-22.
HARRELL, J. C., DYE, W. W., HARVELL, D. M., SARTORIUS, C. A. & HORWITZ, K. B. 2007. Contaminating cells alter gene signatures in whole organ versus laser capture microdissected tumors: a comparison of experimental breast cancers and their lymph node metastases. *Clin Exp Metastasis*.
HARVEY, J. M., DE KLERK, N. H. & STERRETT, G. F. 1992. Histological grading in breast cancer: interobserver agreement, and relation to other prognostic factors including ploidy. *Pathology,* 24**,** 63-8.
HAUSER-KRONBERGER, C. & DANDACHI, N. 2004. Comparison of chromogenic in situ hybridization with other methodologies for HER2 status assessment in breast cancer. *J Mol Histol,* 35**,** 647-53.
HAYES, D. F., ISAACS, C. & STEARNS, V. 2001. Prognostic factors in breast cancer: current and new predictors of metastasis. *J Mammary Gland Biol Neoplasia,* 6**,** 375-92.
HENRIKSEN, K. L., RASMUSSEN, B. B., LYKKESFELDT, A. E., MOLLER, S., EJLERTSEN, B. & MOURIDSEN, H. T. 2007. Semi-quantitative scoring of potentially predictive markers for endocrine treatment of breast cancer: a comparison between whole sections and tissue microarrays. *J Clin Pathol,* 60**,** 397-404.
HERBERT, G. S., SOHN, V. Y. & BROWN, T. A. 2007. The impact of nodal isolated tumor cells on survival of breast cancer patients. *The American Journal of Surgery,* 193**,** 571-574.

HERRINGTON, C. S., LEEK, R. D. & MCGEE, J. O. 1995. Correlation of numerical chromosome 11 and 17 imbalance with metastasis of primary breast cancer to lymph nodes. *J Pathol,* 176**,** 353-9.

HINESTROSA, M. C., DICKERSIN, K., KLEIN, P., MAYER, M., NOSS, K., SLAMON, D., SLEDGE, G. & VISCO, F. M. 2007. Shaping the future of biomarker research in breast cancer to ensure clinical relevance. *Nat Rev Cancer,* 7**,** 309-15.

HIRAKAWA, S., BROWN, L. F., KODAMA, S., PAAVONEN, K., ALITALO, K. & DETMAR, M. 2007. VEGF-C-induced lymphangiogenesis in sentinel lymph nodes promotes tumor metastasis to distant sites. *Blood,* 109**,** 1010-7.

HOANG, C. D., GUILLAUME, T. J., ENGEL, S. C., TAWFIC, S. H., KRATZKE, R. A. & MADDAUS, M. A. 2005. Analysis of paired primary lung and lymph node tumor cells: a model of metastatic potential by multiple genetic programs. *Cancer Detect Prev,* 29**,** 509-17.

HOAR, F. J., CHAUDHRI, S., WADLEY, M. S. & STONELAKE, P. S. 2003. Co-expression of vascular endothelial growth factor C (VEGF-C) and c-erbB2 in human breast carcinoma. *Eur J Cancer,* 39**,** 1698-703.

HU, Z., FAN, C., LIVASY, C., HE, X., OH, D. S., EWEND, M. G., CAREY, L. A., SUBRAMANIAN, S., WEST, R., IKPATT, F., OLOPADE, O. I., VAN DE RIJN, M. & PEROU, C. M. 2009. A compact VEGF signature associated with distant metastases and poor outcomes. *BMC Med,* 7**,** 9.

HU, Z., FAN, C., OH, D. S., MARRON, J. S., HE, X., QAQISH, B. F., LIVASY, C., CAREY, L. A., REYNOLDS, E., DRESSLER, L., NOBEL, A., PARKER, J., EWEND, M. G., SAWYER, L. R., WU, J., LIU, Y., NANDA, R., TRETIAKOVA, M., RUIZ ORRICO, A., DREHER, D., PALAZZO, J. P., PERREARD, L., NELSON, E., MONE, M., HANSEN, H., MULLINS, M., QUACKENBUSH, J. F., ELLIS, M. J., OLOPADE, O. I., BERNARD, P. S. & PEROU, C. M. 2006. The molecular portraits of breast tumors are conserved across microarray platforms. *BMC Genomics,* 7**,** 96.

HUANG, E., CHENG, S. H., DRESSMAN, H., PITTMAN, J., TSOU, M. H., HORNG, C. F., BILD, A., IVERSEN, E. S., LIAO, M. & CHEN, C. M. 2003. Gene expression predictors of breast cancer outcomes. *The Lancet,* 361**,** 1590-1596.

HUANG, Q., GUMIREDDY, K., SCHRIER, M., LE SAGE, C., NAGEL, R., NAIR, S., EGAN, D. A., LI, A., HUANG, G., KLEIN-SZANTO, A. J., GIMOTTY, P. A., KATSAROS, D., COUKOS, G., ZHANG, L., PURE, E. & AGAMI, R. 2008. The microRNAs miR-373 and miR-520c promote tumour invasion and metastasis. *Nat Cell Biol,* 10**,** 202-10.

HURST, D. R., EDMONDS, M. D. & WELCH, D. R. 2009a. Metastamir: the field of metastasis-regulatory microRNA is spreading. *Cancer Research,* 69**,** 7495-7498.

HURST, D. R., EDMONDS, M. D. & WELCH, D. R. 2009b. Metastamir: The Field of Metastasis-Regulatory microRNA Is Spreading. *Cancer Res,* 69**,** 7495-7498.

HURTEAU, G. J., CARLSON, J. A., SPIVACK, S. D. & BROCK, G. J. 2007. Overexpression of the microRNA hsa-miR-200c leads to reduced expression of transcription factor 8 and increased expression of E-cadherin. *Cancer Res,* 67**,** 7972-6.

HUYNH, K. T. & HOON, D. S. 2012. Epigenetics of regional lymph node metastasis in solid tumors. *Clin Exp Metastasis,* 29**,** 747-56.

IGNATIADIS, M. & SOTIRIOU, C. 2008a. Understanding the molecular basis of histologic grade. *Pathobiology,* 75**,** 104-11.

IGNATIADIS, M. & SOTIRIOU, C. 2008b. Understanding the molecular basis of histologic grade. *Pathobiology,* 75**,** 104-11.

IVSHINA, A. V., GEORGE, J., SENKO, O., MOW, B., PUTTI, T. C., SMEDS, J., LINDAHL, T., PAWITAN, Y., HALL, P., NORDGREN, H., WONG, J. E. L., LIU, E. T., BERGH, J., KUZNETSOV, V. A. & MILLER, L. D. 2006. Genetic Reclassification of Histologic Grade Delineates New Clinical Subtypes of Breast Cancer. *Cancer Res,* 66**,** 10292-10301.

IWASAKI, Y., FUKUTOMI, T., AKASHI-TANAKA, S., NANASAWA, T. & TSUDA, H. 1998. Axillary node metastasis from T1N0M0 breast cancer: possible avoidance of dissection in a subgroup. *Jpn J Clin Oncol,* 28**,** 601-3.

JAMARANI, S. M. H., REZAI-RAD, G. & BEHNAM, H. A Novel Method for Breast Cancer Prognosis Using Wavelet Packet Based Neural Network. Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the, 2005. 3414-3417.

JANI, A. B., BASU, A., HEIMANN, R. & HELLMAN, S. 2003. Sentinel lymph node versus axillary lymph node dissection for early-stage breast carcinoma: A comparison using a utility-adjusted number needed to treat analysis. *Cancer,* 97**,** 359-366.

JARMAN, I. H., ETCHELLS, T. A., MARTÍN, J. D. & LISBOA, P. J. G. 2008. An integrated framework for risk profiling of breast cancer patients following surgery. *Artificial Intelligence in Medicine,* 42**,** 165-188.

JARRARD, D. F., KINOSHITA, H., SHI, Y., SANDEFUR, C., HOFF, D., MEISNER, L. F., CHANG, C., HERMAN, J. G., ISAACS, W. B. & NASSIF, N. 1998. Methylation of the androgen receptor promoter

CpG island is associated with loss of androgen receptor expression in prostate cancer cells. *Cancer Res,* 58**,** 5310-4.

JEONG, H. S., JONES, D., LIAO, S., WATTSON, D. A., CUI, C. H., DUDA, D. G., WILLETT, C. G., JAIN, R. K. & PADERA, T. P. 2015. Investigation of the Lack of Angiogenesis in the Formation of Lymph Node Metastases. *J Natl Cancer Inst,* 107.

JEREVALL, P. L., MA, X. J., LI, H., SALUNGA, R., KESTY, N. C., ERLANDER, M. G., SGROI, D. C., HOLMLUND, B., SKOOG, L., FORNANDER, T., NORDENSKJOLD, B. & STAL, O. 2011. Prognostic utility of HOXB13:IL17BR and molecular grade index in early-stage breast cancer patients from the Stockholm trial. *Br J Cancer,* 104**,** 1762-9.

JEREZ, J. M., FRANCO, L., ALBA, E., LLOMBART-CUSSAC, A., LLUCH, A., RIBELLES, N., MUNÃ¡RRIZ, B. & MARTÃ-N, M. 2005. Improvement of breast cancer relapse prediction in high risk intervals using artificial neural networks. *Breast Cancer Research and Treatment,* 94**,** 265-272.

JOENSUU, H., ISOLA, J., LUNDIN, M., SALMINEN, T., HOLLI, K., KATAJA, V., PYLKKANEN, L., TURPEENNIEMI-HUJANEN, T., VON SMITTEN, K. & LUNDIN, J. 2003. Amplification of erbB2 and erbB2 expression are superior to estrogen receptor status as risk factors for distant recurrence in pT1N0M0 breast cancer: a nationwide population-based study. *Clin Cancer Res,* 9**,** 923-30.

JORGENSEN, J., PEDERSEN, B. & PEDERSEN, S. 1996. Use of neural networks to diagnose acute myocardial infarction I methodology. *Clinical Chemistry,* 42**,** 604-612.

KAN, T., SHIMADA, Y., SATO, F., ITO, T., KONDO, K., WATANABE, G., MAEDA, M., YAMASAKI, S., MELTZER, S. J. & IMAMURA, M. 2004. Prediction of lymph node metastasis with use of artificial neural networks based on gene expression profiles in esophageal squamous cell carcinoma. *Ann Surg Oncol,* 11**,** 1070-8.

KANG, Y. 2006. New tricks against an old foe: molecular dissection of metastasis tissue tropism in breast cancer. *Breast Dis,* 26**,** 129-38.

KAPUR, U., RUBINAS, T., GHAI, R., SINACORE, J., YAO, K. & RAJAN, P. B. 2007. Prediction of nonsentinel lymph node metastasis in sentinel node-positive breast carcinoma. *Ann Diagn Pathol,* 11**,** 10-2.

KARLSSON, C., BODIN, L., PIEHL-AULIN, K. & KARLSSON, M. G. 2009. Tissue Microarray Validation: A Methodologic Study with Special Reference to Lung Cancer. *Cancer Epidemiology Biomarkers & Prevention,* 18**,** 2014-2021.

KAY, E., O'GRADY, A., MORGAN, J. M., WOZNIAK, S. & JASANI, B. 2004. Use of tissue microarray for interlaboratory validation of HER2 immunocytochemical and FISH testing. *J Clin Pathol,* 57**,** 1140-1144.

KELLY, P. N., DAKIC, A., ADAMS, J. M., NUTT, S. L. & STRASSER, A. 2007. Tumor growth need not be driven by rare cancer stem cells. *Science,* 317**,** 337.

KIMBUNG, S., LOMAN, N. & HEDENFALK, I. 2015. Clinical and molecular complexity of breast cancer metastases. *Semin Cancer Biol*.

KINOSHITA, H., SHI, Y., SANDEFUR, C., MEISNER, L. F., CHANG, C., CHOON, A., REZNIKOFF, C. R., BOVA, G. S., FRIEDL, A. & JARRARD, D. F. 2000. Methylation of the androgen receptor minimal promoter silences transcription in human prostate cancer. *Cancer Res,* 60**,** 3623-30.

KOLLIAS, J., GILL, P. G., CHATTERTON, B., RAYMOND, W. & COLLINS, P. J. 1999. Sentinel node biopsy in breast cancer: recommendations for surgeons, pathologists, nuclear physicians and radiologists in Australia and New Zealand. *Australian and New Zealand Journal of Surgery,* 70**,** 132-136.

KONECNY, G., PAULETTI, G., PEGRAM, M., UNTCH, M., DANDEKAR, S., AGUILAR, Z., WILSON, C., RONG, H. M., BAUERFEIND, I., FELBER, M., WANG, H. J., BERYT, M., SESHADRI, R., HEPP, H. & SLAMON, D. J. 2003. Quantitative association between HER-2/neu and steroid hormone receptors in hormone receptor-positive primary breast cancer. *J Natl Cancer Inst,* 95**,** 142-53.

KONECNY, G. & SLAMON, D. 2002. HER2 testing and correlation with efficacy of Trastuzumab therapy. *Oncology,* 16.

KONONEN, J., BUBENDORF, L., KALLIONIEMI, A., BARLUND, M., SCHRAML, P., LEIGHTON, S., TORHORST, J., MIHATSCH, M. J., SAUTER, G. & KALLIONIEMI, O. P. 1998. Tissue microarrays for high-throughput molecular profiling of tumor specimens. *Nat Med,* 4**,** 844-7.

KORHONEN, J., MARTINMÄKI, P., PIZZI, C., P., R. & UKKONEN, E. 2009. MOODS: fast search for position weight matrix matches in DNA sequences. *Bioinformatics,* 25**,** 3181-3182.

KOUCHOUKOS, N. T., ACKERMAN, L. V. & BUTCHER, H. R., JR. 1967. Prediction of axillary nodal metastases from the morphology of primary mammary carcinomas. Guide to operative therapy. *Cancer,* 20**,** 948-60.

KOZOMARA, A. & GRIFFITHS-JONES, S. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res,* 39**,** D152-7.

KRAG, D., WEAVER, D., ASHIKAGA, T., MOFFAT, F. L., KLIMBERG, V. S., SHRIVER, C., FELDMAN, S., KUSMINSKY, R., GADD, M., KUHN, J., HARLOW, S. & BEITSH, P. 1998. The sentinel node in breast cancer. *The New England Journal of Medicine,* 339**,** 941-946.

KRAG, D. N., WEAVER, D. L., ALEX, J. C. & FAIRBANK, J. T. 1993. Surgical resection and radiolocalization of the sentinel lymph node in breast cancer using a gamma probe. *Surgical oncology,* 2**,** 335-340.

KREUNIN, P., YOO, C., URQUIDI, V., LUBMAN, D. M. & GOODISON, S. 2007. Proteomic profiling identifies breast tumor metastasis-associated factors in an isogenic model. *Proteomics,* 7**,** 299-312.

KRISHNAN, J., KIRKIN, V., STEFFEN, A., HEGEN, M., WEIH, D., TOMAREV, S., WILTING, J. & SLEEMAN, J. P. 2003. Differential in vivo and in vitro expression of vascular endothelial growth factor (VEGF)-C and VEGF-D in tumors and its relationship to lymphatic metastasis in immunocompetent rats. *Cancer Res,* 63**,** 713-22.

KUNKLER, I., PURUSHOTHAM, A. D., MCINTOSH, S. & COOKE, T. 1999. Sentinel lymph node in breast cancer. *The Lancet,* 254**,** 1998-1999.

KUROSUMI, M., SUEMASU, K., TABEI, T., INOUE, K., MATSUMOTO, H., SUGAMATA, N. & HIGASHI, Y. 2001. Relationship between existence of lymphatic invasion in peritumoral breast tissue and presence of axillary lymph node metastasis in invasive ductal carcinoma of the breast. *Oncol Rep,* 8**,** 1051-5.

KUUKASJARVI, T., KARHU, R., TANNER, M., KAHKONEN, M., SCHAFFER, A., NUPPONEN, N., PENNANEN, S., KALLIONIEMI, A., KALLIONIEMI, O. P. & ISOLA, J. 1997. Genetic heterogeneity and clonal evolution underlying development of asynchronous metastasis in human breast cancer. *Cancer Res,* 57**,** 1597-604.

LAAKSO, M., TANNER, M. & ISOLA, J. 2006. Dual-colour chromogenic in situ hybridization for testing of HER-2 oncogene amplification in archival breast tumours. *J Pathol,* 210**,** 3-9.

LAMBERT, L. A., AYERS, G. D. & MERIC-BERNSTAM, F. 2007. Validation of a breast cancer nomogram for predicting nonsentinel lymph node metastases after a positive sentinel node biopsy. *Ann Surg Oncol,* 14**,** 2422-3.

LANCASHIRE, L. J., REES, R. C. & BALL, G. R. 2008. Identification of gene transcript signatures predictive for estrogen receptor and lymph node status using a stepwise forward selection artificial neural network modelling approach. *Artificial Intelligence in Medicine,* 43**,** 99-111.

LAND, C. E., TOKUNAGA, M., KOYAMA, K., SODA, M., PRESTON, D. L., NISHIMORI, I. & TOKUOKA, S. 2003. Incidence of female breast cancer among atomic bomb survivors, Hiroshima and Nagasaki, 1950-1990. *Radiat Res,* 160**,** 707-17.

LANGER, I., GULLER, U., BERCLAZ, G., KOECHLI, O. R., SCHAER, G., FEHR, M. K., HESS, T., OERTLI, D., BRONZ, L., SCHNARWYLER, B., WIGHT, E., UEHLINGER, U., INFANGER, E., BURGER, D. & ZUBER, M. 2007. Morbidity of sentinel lymph node biopsy (SLN) alone versus SLN and completion axillary lymph node dissection after breast cancer surgery: a prospective Swiss multicenter study on 659 patients. *Ann Surg,* 245**,** 452-61.

LEE, H., LIN, E. C., LIU, L. & SMITH, J. W. 2003. Gene expression profiling of tumor xenografts: In vivo analysis of organ-specific metastasis. *Int J Cancer,* 107**,** 528-34.

LEWIS, B. P., BURGE, C. B. & BARTEL, D. P. 2005. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell,* 120**,** 15-20.

LI, J., GROMOV, P., GROMOVA, I., MOREIRA, J. M., TIMMERMANS-WIELENGA, V., RANK, F., WANG, K., LI, S., LI, H., WIUF, C., YANG, H., ZHANG, X., BOLUND, L. & CELIS, J. E. 2008. Omics-based profiling of carcinoma of the breast and matched regional lymph node metastasis. *Proteomics,* 8**,** 5038-52.

LI, S., HAN, B., LIU, G., OUELLET, J., LABRIE, F. & PELLETIER, G. 2010. Immunocytochemical localization of sex steroid hormone receptors in normal human mammary gland. *J Histochem Cytochem,* 58**,** 509-15.

LI, Y.-S., KANEKO, M., AMATYA, V. J., TAKESHIMA, Y., ARIHIRO, K. & INAI, K. 2006. Expression of vascular endothelial growth factor-C and its receptor in invasive micropapillary carcinoma of the breast. *Pathology International,* 56**,** 256-261.

LI, Y., ROGOFF, H. A., KEATES, S., GAO, Y., MURIKIPUDI, S., MIKULE, K., LEGGETT, D., LI, W., PARDEE, A. B. & LI, C. J. 2015. Suppression of cancer relapse and metastasis by inhibiting cancer stemness. *Proc Natl Acad Sci U S A,* 112**,** 1839-44.

LIAO, D. J. & DICKSON, R. B. 2002. Roles of androgens in the development, growth, and carcinogenesis of the mammary gland. *J Steroid Biochem Mol Biol,* 80**,** 175-89.

LISBOA, P. J. G., ETCHELLS, T. A., JARMAN, I. H., AUNG, M. S. H., CHABAUD, S., BACHELOR, T., PEROL, D., GARGI, T., BOURDES, V., BONNEVAY, S. & NEGRIER, S. Time-to-event analysis with artificial neural networks: An integrated analytical and rule-based study for breast cancer. Neural Networks, 2007. IJCNN 2007. International Joint Conference on, 2007. 2533-2538.

LISBOA, P. J. G., ETCHELLS, T. A., JARMAN, I. H., HANE AUNG, M. S., CHABAUD, S., BACHELOT, T., PEROL, D., GARGI, T., BOURDÈS, V., BONNEVAY, S. & NÉGRIER, S. 2008. Time-to-event analysis with artificial neural networks: An integrated analytical and rule-based study for breast cancer. *Neural Networks,* 21**,** 414-426.

LIU, F., LANG, R., WEI, J., FAN, Y., CUI, L., GU, F., GUO, X., PRINGLE, G. A., ZHANG, X. & FU, L. 2009. Increased expression of SDF-1/CXCR4 is associated with lymph node metastasis of invasive micropapillary carcinoma of the breast. *Histopathology,* 54**,** 741-50.

LIU, R., WANG, X., CHEN, G. Y., DALERBA, P., GURNEY, A., HOEY, T., SHERLOCK, G., LEWICKI, J., SHEDDEN, K. & CLARKE, M. F. 2007. The Prognostic Role of a Gene Signature from Tumorigenic Breast-Cancer Cells. *N Engl J Med,* 356**,** 217-226.

LUNDIN, J., LUNDIN, M., HOLLI, K., KATAJA, V., ELOMAA, L., PYLKKANEN, L., TURPEENNIEMI-HUJANEN, T. & JOENSUU, H. 2001. Omission of histologic grading from clinical decision making may result in overuse of adjuvant therapies in breast cancer: results from a nationwide study. *J Clin Oncol,* 19**,** 28-36.

LUNDIN, M., LUNDIN, J., BURKE, H. B., TOIKKANEN, S., PYLKKANEN, L. & JOENSUU, H. 1999. Artificial neural networks applied to survival prediction in breast cancer. *Oncology,* 57**,** 281-6.

LUO, X., SHI, Y. X., LI, Z. M. & JIANG, W. Q. 2010. Expression and clinical significance of androgen receptor in triple negative breast cancer. *Chin J Cancer,* 29**,** 585-90.

MA, L., TERUYA-FELDSTEIN, J. & WEINBERG, R. A. 2007a. Tumour invasion and metastasis initiated by microRNA-10b in breast cancer. *Nature,* 449**,** 682-8.

MA, X.-J., SALUNGA, R., DAHIYA, S., WANG, W., CARNEY, E., DURBECQ, V., HARRIS, A., GOSS, P., SOTIRIOU, C., ERLANDER, M. & SGROI, D. 2008. A Five-Gene Molecular Grade Index and HOXB13:IL17BR Are Complementary Prognostic Factors in Early Stage Breast Cancer. *Clin Cancer Res,* 14**,** 2601-2608.

MA, X., SALUNGA, R., TUGGLE, J. T., GAUDET, J., ENRIGHT, E., MCQUARY, P., PAYETTE, T., PISTONE, M., STECKER, K., ZHANG, B. M., ZHOU, Y., VARNHOLT, H., SMITH, B., GADD, M., CHATFIELD, E., KESSLER, J., BAER, T. M. & SGROI, M. G. E. D. C. 2003. Gene expression profiles of human breast cancer progression. *Proc Natl Acad Sci,* 100**,** 5974-5979.

MA, Y., QIAN, Y., WEI, L., ABRAHAM, J., SHI, X., CASTRANOVA, V., HARNER, E. J., FLYNN, D. C. & GUO, L. 2007b. Population-Based Molecular Prognosis of Breast Cancer by Transcriptional Profiling. *Clin Cancer Res,* 13**,** 2014-2022.

MACKILN, P. S., DEMPSY, J., BROOKS, J. & RAND, J. 1991. Using Neural Networks to Diagnose Cancer. *Journal of Medical Systems,* 15**,** 11-19.

MAIBENCO, D. C., DOMBI, G. W., KAU, T. Y. & SEVERSON, R. K. 2006. Significance of micrometastases on the survival of women with T1 breast cancer. *Cancer,* 107**,** 1234-1239.

MALKAS, L. H., HERBERT, B. S., ABDEL-AZIZ, W., DOBROLECKI, L. E., LIU, Y., AGARWAL, B., HOELZ, D., BADVE, S., SCHNAPER, L., ARNOLD, R. J., MECHREF, Y., NOVOTNY, M. V., LOEHRER, P., GOULET, R. J. & HICKEY, R. J. 2006. A cancer-associated PCNA expressed in breast cancer has implications as a potential biomarker. *PNAS***,** 0604614103.

MANNWEILER, S., TSYBROVSKYY, O. & REGAUER, S. 2002. The flow cytometric DNA index can predict the presence of lymph node metastases in invasive ductal breast carcinoma. *Apmis,* 110**,** 580-6.

MANSEL, R. E., FALLOWFIELD, L., KISSIN, M., GOYAL, A., NEWCOMBE, R. G., DIXON, J. M., YIANGOU, C., HORGAN, K., BUNDRED, N., MONYPENNY, I., ENGLAND, D., SIBBERING, M., ABDULLAH, T. I., BARR, L., CHETTY, U., SINNETT, D. H., FLEISSIG, A., CLARKE, D. & ELL, P. J. 2006. Randomized multicenter trial of sentinel node biopsy versus standard axillary treatment in operable breast cancer: the ALMANAC Trial. *J Natl Cancer Inst,* 98**,** 599-609.

MANSI, J. L., GOGAS, H., BLISS, J. M., GAZET, J. C., BERGER, U. & COOMBES, R. C. 1999. Outcome of primary-breast-cancer patients with micrometastases: a long-term follow-up study. *Lancet,* 354**,** 197-202.

MANZOTTI, M., DELL'ORTO, P., MAISONNEUVE, P., ZURRIDA, S., MAZZAROL, G. & VIALE, G. 2001. Reverse transcription-polymerase chain reaction assay for multiple mRNA markers in the detection of breast cancer metastases in sentinel lymph nodes. *Int J Cancer,* 95**,** 307-12.

MARCHEVSKY, A. M., PATEL, S., WILEY, K. J., STEPHENSON, M. A., GONDO, M., BROWN, R. W., YI, E. S., BENEDICT, W. F., ANTON, R. C. & CAGLE, P. T. 1998. Artifical Neural Networks and Logistic Regression as Tools for prediction of survival in patients with stages I and II Non-small cell lung cancer. *Mod Pathol,* 11**,** 618-628.

MARCHEVSKY, A. M., SHAH, S. & PATEL, S. 1999. Reasoning with uncertainty in pathology: artificial neural networks and logistic regression as tools for prediction of lymph node status in breast cancer patients. *Mod Pathol,* 12**,** 505-13.

MARINHO, V. F., METZE, K., SANCHES, F. S., ROCHA, G. F. & GOBBI, H. 2008. Lymph vascular invasion in invasive mammary carcinomas identified by the endothelial lymphatic marker D2-40 is associated with other indicators of poor prognosis. *BMC Cancer,* 8**,** 64.

MARINHO, V. F., ZAGURY, M. S., CALDEIRA, L. G. & GOBBI, H. 2006. Relationship between histologic features of primary breast carcinomas and axillary lymph node micrometastases: Detection and prognostic significance. *Appl Immunohistochem Mol Morphol,* 14**,** 426-31.

MARKIEWICZ, A., WELNICKA-JASKIEWICZ, M., SEROCZYNSKA, B., SKOKOWSKI, J., MAJEWSKA, H., SZADE, J. & ZACZEK, A. J. 2014. Epithelial-mesenchymal transition markers in lymph node metastases and primary breast tumors - relation to dissemination and proliferation. *Am J Transl Res,* 6**,** 793-808.

MASOOD, S. & BUI, M. M. 2002. Prognostic and predictive value of HER2/neu oncogene in breast cancer. *Microscopy Research and Technique,* 59**,** 102-108.

MASS, R. D., PRESS, M. F., ANDERSON, S., COBLEIGH, M. A., VOGEL, C. L., DYBDAL, N., LEIBERMAN, G. & SLAMON, D. J. 2005. Evaluation of clinical outcomes according to HER2 detection by fluorescence in situ hybridization in women with metastatic breast cancer treated with trastuzumab. *Clin Breast Cancer,* 6**,** 240-6.

MASTERS, K. M., GIULIANO, A. E., ROSS, M. I., REINTGEN, D. S., HUNT, K. K., BYRD, D. R., KLIMBERG, V. S., WHITWORTH, P. W., TAFRA, L. C. & EDWARDS, M. J. 1998. Sentinel-Lymph node biopsy for breast cancer - not yet the standard of care. *The New England Journal of Medicine,* 339**,** 990-995.

MATTFELDT, T., KESTLER, H. A. & SINN, H. P. 2004. Prediction of the axillary lymph node status in mammary cancer on the basis of clinicopathological data and flow cytometry. *Med Biol Eng Comput,* 42**,** 733-9.

MCGUIRE, A., BROWN, J. A. & KERIN, M. J. 2015. Metastatic breast cancer: the potential of miRNA for diagnosis and treatment monitoring. *Cancer Metastasis Rev,* 34**,** 145-55.

MCGUIRE, W. L. 1987. Prognostic factors for recurrence and survival in human breast cancer. *Breast Cancer Res Treat,* 10**,** 5-9.

MCGUIRE, W. L. & CLARK, G. M. 1992. Prognostic factors and treatment decisions in axillary-node-negative breast cancer. *N Engl J Med,* 326**,** 1756-61.

MCNAMARA, K. M., MOORE, N. L., HICKEY, T. E., SASANO, H. & TILLEY, W. D. 2014. Complexities of androgen receptor signalling in breast cancer. *Endocr Relat Cancer,* 21**,** T161-81.

MELCHERS, L. J., CLAUSEN, M., MASTIK, M. F., SLAGTER-MENKEMA, L., VAN DER WAL, J. E., WISMAN, G., ROODENBURG, J. & SCHUURING, E. 2015. Identification of methylation markers for the prediction of nodal metastasis in oral and oropharyngeal squamous cell carcinoma. *Epigenetics,* 10**,** 850-60.

MENARD, S., CASCINELLI, N., RILKE, F. & COLNAGHI, M. I. 1995. Re: Prediction of axillary lymph node status in breast cancer patients by use of prognostic indicators. *J Natl Cancer Inst,* 87**,** 607-8.

MENGEL, M., KREIPE, H. & VON WASIELEWSKI, R. 2003. Rapid and large-scale transition of new tumor biomarkers to clinical biopsy material by innovative tissue microarray systems. *Appl Immunohistochem Mol Morphol,* 11**,** 261-8.

MICHOR, F., NOWAK, M. A. & IWASA, Y. 2006. Stochastic dynamics of metastasis formation. *J Theor Biol,* 240**,** 521-30.

MIRANDA, K. C., HUYNH, T., TAY, Y., ANG, Y. S., TAM, W. L., THOMSON, A. M., LIM, B. & RIGOUTSOS, I. 2006. A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes. *Cell,* 126**,** 1203-17.

MITTENDORF, E. A. & HUNT, K. K. 2007. Significance and management of micrometastases in patients with breast cancer. *Expert Rev Anticancer Ther,* 7**,** 1451-61.

MIZOKAMI, A., YEH, S. Y. & CHANG, C. 1994. Identification of 3',5'-cyclic adenosine monophosphate response element and other cis-acting elements in the human androgen receptor gene promoter. *Mol Endocrinol,* 8**,** 77-88.

MOBASHERI, A., AIRLEY, R., FOSTER, C. S., SCHULZE-TANZIL, G. & SHAKIBAEI, M. 2004. Post-genomic applications of tissue microarrays: basic research, prognostic oncology, clinical genomics and drug discovery. *Histol Histopathol,* 19**,** 325-35.

MOFFAT, F. L., JR. 2001. Sentinel node biopsy is not an alternative to axillary dissection in breast cancer. *J Surg Oncol,* 77**,** 153-6.

MOHAMMED, R. A., GREEN, A., EL-SHIKH, S., PAISH, E. C., ELLIS, I. O. & MARTIN, S. G. 2007a. Prognostic significance of vascular endothelial cell growth factors -A, -C and -D in breast cancer and their relationship with angio- and lymphangiogenesis. *Br J Cancer,* 96**,** 1092-1100.

MOHAMMED, R. A., MARTIN, S. G., GILL, M. S., GREEN, A. R., PAISH, E. C. & ELLIS, I. O. 2007b. Improved Methods of Detection of Lymphovascular Invasion Demonstrate That It is the Predominant

Method of Vascular Invasion in Breast Cancer and has Important Clinical Consequences. *Am J Surg Pathol,* 31**,** 1825-1833.

MONTEL, V., HUANG, T. Y., MOSE, E., PESTONJAMASP, K. & TARIN, D. 2005. Expression profiling of primary tumors and matched lymphatic and lung metastases in a xenogeneic breast cancer model. *Am J Pathol,* 166**,** 1565-79.

MUMPRECHT, V. & DETMAR, M. 2009. Lymphangiogenesis and Cancer Metastasis. *J Cell Mol Med*.

MUSSURAKIS, S., BUCKLEY, D. L. & HORSMAN, A. 1997. Prediction of axillary lymph node status in invasive breast cancer with dynamic contrast-enhanced MR imaging. *Radiology,* 203**,** 317-21.

NAGUIB, R. N., SAKIM, H. A., LAKSHMI, M. S., WADEHRA, V., LENNARD, T. W., BHATAVDEKAR, J. & SHERBET, G. V. 1999. DNA ploidy and cell cycle distribution of breast cancer aspirate cells measured by image cytometry and analyzed by artificial neural networks for their prognostic significance. *IEEE Trans Inf Technol Biomed,* 3**,** 61-9.

NAGUIB, R. N. G., ADAMS, A. E., HORNE, C. H. W., ANGUS, B., SMITH, A. F., SHERBET, G. V. & LENNARD, T. W. J. 1997. Prediction of Nodal Metastasis and Prognosis in Breast Cancer: A Neural Mode. *Anticancer Research,* 17**,** 2735-2742.

NAGUIB, R. N. G. & SHERBERT, G. V. 1997. Artificial Neural Networks in Cancer Research. *Pathobiology,* 65**,** 129-139.

NAKAGAWA, T., HUANG, S. K., MARTINEZ, S. R., TRAN, A. N., ELASHOFF, D., YE, X., TURNER, R. R., GIULIANO, A. E. & HOON, D. S. B. 2006. Proteomic Profiling of Primary Breast Cancer Predicts Axillary Lymph Node Metastasis. *Cancer Res,* 66**,** 11825-11830.

NAKAMURA, K., YAMASHITA, K., SAWAKI, H., WARAYA, M., KATOH, H., NAKAYAMA, N., KAWAMATA, H., NISHIMIYA, H., EMA, A., NARIMATSU, H. & WATANABE, M. 2015. Aberrant methylation of GCNT2 is tightly related to lymph node metastasis of primary CRC. *Anticancer Res,* 35**,** 1411-21.

NAKAMURA, Y., YASUOKA, H., TSUJIMOTO, M., IMABUN, S., NAKAHARA, M., NAKAO, K., NAKAMURA, M., MORI, I. & KAKUDO, K. 2005. Lymph vessel density correlates with nodal status, VEGF-C expression, and prognosis in breast cancer. *Breast Cancer Res Treat,* 91**,** 125-32.

NAKAMURA, Y., YASUOKA, H., TSUJIMOTO, M., YANG, Q., IMABUN, S., NAKAHARA, M., NAKAO, K., NAKAMURA, M., MORI, I. & KAKUDO, K. 2003. Flt-4-positive vessel density correlates with vascular endothelial growth factor-d expression, nodal status, and prognosis in breast cancer. *Clin Cancer Res,* 9**,** 5313-7.

NAKAMURA, Y., YASUOKA, H., TSUJIMOTO, M., YOSHIDOME, K., NAKAHARA, M., NAKAO, K., NAKAMURA, M. & KAKUDO, K. 2006. Nitric Oxide in Breast Cancer: Induction of Vascular Endothelial Growth Factor-C and Correlation with Metastasis and Poor Prognosis. *Clin Cancer Res,* 12**,** 1201-1207.

NAKOPOULOU, L., PANAYOTOPOULOU, E. G., GIANNOPOULOU, I., TSIRMPA, I., KATSAROU, S., MYLONA, E., ALEXANDROU, P. & KERAMOPOULOS, A. 2007. Extra copies of chromosomes 16 and X in invasive breast carcinomas are related to aggressive phenotype and poor prognosis. *J Clin Pathol,* 60**,** 808-815.

NATHANSON, S. D., SLATER, R., DEBRUYN, D., KAPKE, A. & LINDEN, M. 2006. Her-2/neu expression in primary breast cancer with sentinel lymph node metastasis. *Ann Surg Oncol,* 13**,** 205-13.

NI, M., ., CHEN, Y., LIM, E., WIMBERLY, H., BAILEY, S. T., IMAI, Y., RIMM, D. L., LIU, X. S. & BROWN, M. 2011. Targeting androgen receptor in estrogen receptor-negative breast cancer. *Cancer Cell,* 20**,** 119-131.

NICHOLAS, E. R. 1997. Histologic grade as a prognostic factor in breast carcinoma--reply. *Cancer,* 80**,** 1706-1707.

NIEWEG, O. E., VRIES, J. D., JANSEN, L., KOOPS, S. & KROON, B. B. R. 1997. Sentinel-node biopsy in breast cancer. *The Lancet,* 350**,** 808.

NISHIZAKI, T., DEVRIES, S., CHEW, K., GOODSON, W. H., 3RD, LJUNG, B. M., THOR, A. & WALDMAN, F. M. 1997. Genetic alterations in primary breast cancers and their metastases: direct comparison using modified comparative genomic hybridization. *Genes Chromosomes Cancer,* 19**,** 267-72.

NOGUCHI, M., THOMAS, M., KITAGAWA, H., KINOSHITA, K., OHTA, N., NAGAMORI, M. & MIYAZAKI, I. 1993. Further analysis of predictive value of Helix pomatia lectin binding to primary breast cancer for axillary and internal mammary lymph node metastases. *Br J Cancer,* 67**,** 1368-71.

NOWELL, P. C. 1976. The clonal evolution of tumor cell populations. *Science,* 194**,** 23-8.

O'DONNELL, R. K., KUPFERMAN, M., WEI, S. J., SINGHAL, S., WEBER, R., O'MALLEY, B., CHENG, Y., PUTT, M., FELDMAN, M., ZIOBER, B. & MUSCHEL, R. J. 2005. Gene expression signature predicts lymphatic metastasis in squamous cell carcinoma of the oral cavity. *Oncogene,* 24**,** 1244-51.

O'GRADY, A., FLAHAVAN, C., KAY, E. W., BARRETT, H. L. & LEADER, M. B. 2003. HER-2 analysis in tissue microarrays of archival human breast cancer: Comarison of immunohistochemistry and

fluorescence in situ hybridization. *Applied Immunohistochemistry and Molecular Morphology,* 11**,** 177-182.

OLIVOTTO, I. A., JACKSON, J. S., MATES, D., ANDERSEN, S., DAVIDSON, W., BRYCE, C. J. & RAGAZ, J. 1998. Prediction of axillary lymph node involvement of women with invasive breast carcinoma: a multivariate analysis. *Cancer,* 83**,** 948-55.

OZMEN, V. & CABIOGLU, N. 2006. Sentinel lymph node biopsy for breast cancer: current controversies. *Breast J,* 12**,** S134-42.

PADERA, T. P., KADAMBI, A., DI TOMASO, E., CARREIRA, C. M., BROWN, E. B., BOUCHER, Y., CHOI, N. C., MATHISEN, D., WAIN, J., MARK, E. J., MUNN, L. L. & JAIN, R. K. 2002. Lymphatic metastasis in the absence of functional intratumor lymphatics. *Science,* 296**,** 1883-6.

PAIK, S., BRYANT, J., PARK, C., FISHER, B., TAN-CHIU, E., HYAMS, D., FISHER, E. R., LIPPMAN, M. E., WICKERHAM, D. L. & WOLMARK, N. 1998. erb-2 and response to Doxorubicin in patients with axillary lymph node-positive, hormone receptor-negative breas cancer. *Journal of the National Cancer Institute,* 90**,** 1361-1370.

PAIK, S. & PARK, C. 2001. HER-2 and choice of adjuvant chemotherapy in breast cancer. *Seminars in Oncology,* 28**,** 332-335.

PAL, A., PROVENZANO, E., DUFFY, S. W., PINDER, S. E. & PURUSHOTHAM, A. D. 2007. A model for predicting non-sentinel lymph node metastatic disease when the sentinel lymph node is positive. *Br J Surg*.

PANDIS, N., HEIM, S., BARDI, G., IDVALL, I., MANDAHL, N. & MITELMAN, F. 1993. Chromosome analysis of 20 breast carcinomas: cytogenetic multiclonality and karyotypic-pathologic correlations. *Genes Chromosomes Cancer,* 6**,** 51-7.

PANDIS, N., IDVALL, I., BARDI, G., JIN, Y., GORUNOVA, L., MERTENS, F., OLSSON, H., INGVAR, C., BEROUKAS, K., MITELMAN, F. & HEIM, S. 1996. Correlation between karyotypic pattern and clincopathologic features in 125 breast cancer cases. *Int J Cancer,* 66**,** 191-6.

PARK, K., KIM, J., LIM, S., HAN, S. & LEE, J. Y. 2003. Comparing fluorescence in situ hybridization and chromogenic in situ hybridization methods to determine the HER2/neu status in primary breast carcinoma using tissue microarray. *Mod Pathol,* 16**,** 937-43.

PARK, S., KOO, J. S., KIM, M. S., PARK, H. S., LEE, J. S., KIM, S. I., PARK, B. W. & LEE, K. S. 2011. Androgen receptor expression is significantly associated with better outcomes in estrogen receptor-positive breast cancers. *Ann Oncol*.

PARK, S. Y., KIM, B. H., KIM, J. H., LEE, S. & KANG, G. H. 2007. Panels of immunohistochemical markers help determine primary sites of metastatic adenocarcinoma. *Arch Pathol Lab Med,* 131**,** 1561-7.

PARKER, J. S., MULLINS, M., CHEANG, M. C., LEUNG, S., VODUC, D., VICKERY, T., DAVIES, S., FAURON, C., HE, X., HU, Z., QUACKENBUSH, J. F., STIJLEMAN, I. J., PALAZZO, J., MARRON, J. S., NOBEL, A. B., MARDIS, E., NIELSEN, T. O., ELLIS, M. J., PEROU, C. M. & BERNARD, P. S. 2009. Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol,* 27**,** 1160-7.

PATANI, N. R., DWEK, M. V. & DOUEK, M. 2007. Predictors of axillary lymph node metastasis in breast cancer: a systematic review. *Eur J Surg Oncol,* 33**,** 409-19.

PATHOLOGISTS, C. O. A. 2002. Clinical laboratory assays for HER-2/neu amplification and overexpression: Quality assurance, standardization, and proficiency testing. *Arch Pathol Lab Med,* 126**,** 803-808.

PEDERSEN, S. M., JORGENSEN, J. S. & PEDERSEN, J. B. 1996. Use of neural networks to diagnose acute myocardial infarction. II. A clinical application. *Clinical Chemistry,* 42**,** 613-617.

PEIRO, G., ADROVER, E., ARANDA, F. I., PEIRO, F. M., NIVEIRO, M. & SANCHEZ-PAYA, J. 2007a. Prognostic implications of HER-2 status in steroid receptor-positive, lymph node-negative breast carcinoma. *Am J Clin Pathol,* 127**,** 780-6.

PEIRO, G., ARANDA, F. I., ADROVER, E., NIVEIRO, M., ALENDA, C., PAYA, A. & SEGUI, J. 2007b. Analysis of HER2 by chromogenic in situ hybridization and immunohistochemistry in lymph node-negative breast carcinoma: Prognostic relevance. *Hum Pathol,* 38**,** 26-34.

PEREZ, E. A., ROCHE, P., JENKINS, R., REYNOLDS, C., HALLING, K., INGLE, J. & WOLD, L. 2002. HER2 testing in patients with breast cancer:Poor correlation between weak positivity by immunohistochemistry and gene amplification by fluorescence in situ hybridization. *Mayo Clin Proc,* 77**,** 148-154.

PEROU, C. M., SORLIE, T., EISEN, M. B., VAN DE RIJN, M., JEFFREY, S. S., REES, C. A., POLLACK, J. R., ROSS, D. T., JOHNSEN, H., AKSLEN, L. A., FLUGE, O., PERGAMENSCHIKOV, A., WILLIAMS, C., ZHU, S. X., LONNING, P. E., BORRESEN-DALE, A. L., BROWN, P. O. & BOTSTEIN, D. 2000. Molecular portraits of human breast tumours. *Nature,* 406**,** 747-52.

PETERS, K. M., EDWARDS, S. L., NAIR, S. S., FRENCH, J. D., BAILEY, P. J., SALKIELD, K., STEIN, S., WAGNER, S., FRANCIS, G. D., CLARK, S. J. & BROWN, M. A. 2012. Androgen receptor

expression predicts breast cancer survival: the role of genetic and epigenetic events. *BMC Cancer,* 12**,** 132.

PIAO, Z. & MALKHOSYAN, S. R. 2002. Frequent loss Xq25 on the inactive X chromosome in primary breast carcinomas is associated with tumor grade and axillary lymph node metastasis. *Genes Chromosomes Cancer,* 33**,** 262-9.

PICKREN, J. W. 1961. Significance of occult metastases. A study of breast cancer. *Cancer,* 14**,** 1266-71.

PLANT, H. C., KASHYAP, A. S., MANTON, K. J., HOLLIER, B. G., HURST, C. P., STEIN, S. R., FRANCIS, G. D., BEADLE, G. F., UPTON, Z. & LEAVESLEY, D. I. 2014. Differential subcellular and extracellular localisations of proteins required for insulin-like growth factor- and extracellular matrix-induced signalling events in breast cancer progression. *BMC Cancer,* 14**,** 627.

PONZONE, R., MAGGIOROTTO, F., MARIANI, L., JACOMUZZI, M. E., MAGISTRIS, A., MININANNI, P., BIGLIA, N. & SISMONDI, P. 2007. Comparison of two models for the prediction of nonsentinel node metastases in breast cancer. *Am J Surg,* 193**,** 686-92.

PROMISH, D. I. 1999. Prediction of axillary lymph node involvement of women with invasive breast carcinoma: a multivariate analysis. *Cancer,* 85**,** 1201-3.

QIAN, B., KATSAROS, D., LU, L., PRETI, M., DURANDO, A., ARISIO, R., MU, L. & YU, H. 2009. High miR-21 expression in breast cancer associated with poor disease-free survival in early stage disease and high TGF-beta1. *Breast Cancer Res Treat,* 117**,** 131-40.

QUERZOLI, P., PEDRIALI, M., RINALDI, R., LOMBARDI, A. R., BIGANZOLI, E., BORACCHI, P., FERRETTI, S., FRASSON, C., ZANELLA, C., GHISELLINI, S., AMBROGI, F., ANTOLINI, L., PIANTELLI, M., IACOBELLI, S., MARUBINI, E., ALBERTI, S. & NENCI, I. 2006. Axillary Lymph Node Nanometastases Are Prognostic Factors for Disease-Free Survival and Metastatic Relapse in Breast Cancer Patients. *Clin Cancer Res,* 12**,** 6696-6701.

RAKHA, E. A., REIS-FILHO, J. S., BAEHNER, F., DABBS, D. J., DECKER, T., EUSEBI, V., FOX, S. B., ICHIHARA, S., JACQUEMIER, J., LAKHANI, S. R., PALACIOS, J., RICHARDSON, A. L., SCHNITT, S. J., SCHMITT, F. C., TAN, P. H., TSE, G. M., BADVE, S. & ELLIS, I. O. 2010a. Breast cancer prognostic classification in the molecular era: the role of histological grade. *Breast Cancer Res,* 12**,** 207.

RAKHA, E. A., REIS-FILHO, J. S. & ELLIS, I. O. 2010b. Combinatorial biomarker expression in breast cancer. *Breast Cancer Res Treat,* 120**,** 293-308.

RAMASWAMY, S., ROSS, K. N., LANDER, E. S. & GOLUB, T. R. 2003. A molecular signature of metastasis in primary solid tumors. *Nat Genet,* 33**,** 49-54.

RAVDIN, P. M. & CLARK, G. M. 1992. A practical application of neural network analysis for predicting outcome of individual breast cancer patients. *Breast Cancer Research and Treatment,* 22**,** 285-293.

RAVDIN, P. M., CLARK, G. M., HILSENBECK, S. G., OWENS, M. A., VENDELY, P., PANDIN, M. R. & MCGUIRE, W. L. 1992. A demonstration that breast cancer recurrence can be predicted by neural network analysis. *Breast Cancer Research and Treatment,* 21**,** 47-53.

RAVDIN, P. M., DE LAURENTIIS, M., VENDELY, T. & CLARK, G. M. 1994. Prediction of axillary lymph node status in breast cancer patients by use of prognostic indicators. *J Natl Cancer Inst,* 86**,** 1771-5.

RECAMIER, J. 1829. Recherches sur la Traitment du Cancer sur la Compression Methodique Simple ou Combinee et sur l'Histoire Generale de la Meme Maladie. *Recherches sur le traitment du cancer.* Paris: Gabon.

REVILLION, F., LHOTELLIER, V., HORNEZ, L., LEROY, A., BARANZELLI, M. C., GIARD, S., BONNETERRE, J. & PEYRAT, J. P. 2008. Real-time reverse-transcription PCR to quantify a panel of 19 genes in breast cancer: relationships with sentinel lymph node invasion. *Int J Biol Markers,* 23**,** 10-7.

RIOU, G., MATHIEU, M.-C., BARROIS, M., BIHAN, M.-L. L., AHOMADEGBE, J.-C., BENARD, J. & LE, M. G. 2001. c-erb-2 (HER2/neu) gene amplification is a better indicator of poor prognosis than protein over-expression in operable breast-cancer patients. *Int J Cancer (Pred Oncol),* 95**,** 266-270.

ROBBINS, P., PINDER, S., DE KLERK, N., DAWKINS, H., HARVEY, J., STERRETT, G., ELLIS, I. & ELSTON, C. 1995. Histological grading of breast carcinomas: a study of interobserver agreement. *Hum Pathol,* 26**,** 873-9.

ROBERTI, N. E. 1997. The role of histologic grading in the prognosis of patients with carcinoma of the breast: is this a neglected opportunity? *Cancer,* 80**,** 1708-16.

ROBERTS, N., KLOOS, B., CASSELLA, M., PODGRABINSKA, S., PERSAUD, K., WU, Y., PYTOWSKI, B. & SKOBE, M. 2006. Inhibition of VEGFR-3 activation with the antagonistic antibody more potently suppresses lymph node and distant metastases than inactivation of VEGFR-2. *Cancer Res,* 66**,** 2650-7.

ROSEN, P. P., SAIGO, P. E., BRAUN, D. W., JR., WEATHERS, E. & DEPALO, A. 1981. Predictors of recurrence in stage I (T1N0M0) breast carcinoma. *Ann Surg,* 193**,** 15-25.

ROVERE, G. Q. D. & BIRD, P. A. 1998. Sentinel-lymph-node biopsy breast cancer. *The Lancet,* 352**,** 421-422.

ROZENBERG, S., LIEBENS, F. & HAM, H. 1999. The sentinel node in breast cancer: acceptable false-negative rate. *The Lancet,* 353**,** 1937-1938.

RUI, H. & LEBARON, M. J. 2005. Creating tissue microarrays by cutting-edge matrix assembly. *Expert Rev Med Devices,* 2**,** 673-80.

RUIZ, C., SEIBT, S., KURAYA, K. A., SIRAJ, A. K., MIRLACHER, M., SCHRAML, P., MAURER, R., SPICHTIN, H., TORHORST, J., POPOVSKA, S., SIMON, R. & SAUTER, G. 2005. Tissue microarrays for comparing molecular features with proliferation activity in breast cancer. *Int J Cancer*.

RUSNAK, D. W., AFFLECK, K., COCKERILL, S. G., STUBBERFIELD, C., HARRIS, R., PAGE, M., SMITH, K. J., GUNTRIP, S. B., CARTER, M. C., SHAW, R. J., JOWETT, A., STABLES, J., TOPLEY, P., WOOD, E. R., BRIGNOLA, P. S., KADWELL, S. H., REEP, B. R., MULLIN, R. J., ALLIGOOD, K. J., KEITH, B. R., CROSBY, R. M., MURRAY, D. M., KNIGHT, W. B., GILMER, T. M. & LACKEY, K. 2001. The characterization of novel, dual ErbB-2/EGFR, tyrosine kinase inhibitors: potential therapy for cancer. *Cancer Res,* 61**,** 7196-203.

RYDEN, L., LANDBERG, G., STAL, O., NORDENSKJOLD, B., FERNO, M. & BENDAHL, P. O. 2007. HER2 status in hormone receptor positive premenopausal primary breast cancer adds prognostic, but not tamoxifen treatment predictive, information. *Breast Cancer Res Treat*.

S.-J. KUO, Y. H. H. Y. L. H. D. R. C. 2008. Classification of benign and malignant breast tumors using neural networks and three-dimensional power Doppler ultrasound. *Ultrasound in Obstetrics and Gynecology,* 32**,** 97-102.

SAKAGUCHI, M., VIRMANI, A., DUDAK, M. W., PETERS, G. N., LEITCH, A. M., SABOORIAN, H., GAZDAR, A. F. & EUHUS, D. M. 2003. Clinical relevance of reverse transcriptase-polymerase chain reaction for the detection of axillary lymph node metastases in breast cancer. *Ann Surg Oncol,* 10**,** 117-25.

SANTINELLI, A., PISA, E., STRAMAZZOTTI, D. & FABRIS, G. 2008. HER-2 status discrepancy between primary breast cancer and metastatic sites. Impact on target therapy. *Int J Cancer,* 122**,** 999-1004.

SANTOS, S. C., CAVALLI, I. J., RIBEIRO, E. M., URBAN, C. A., LIMA, R. S., BLEGGI-TORRES, L. F., RONE, J. D., HADDAD, B. R. & CAVALLI, L. R. 2008. Patterns of DNA copy number changes in sentinel lymph node breast cancer metastases. *Cytogenet Genome Res,* 122**,** 16-21.

SAYED, D. & ABDELLATIF, M. 2011. MicroRNAs in development and disease. *Physiol Rev,* 91**,** 827-87.

SCHILLACI, O. & SCOPINARO, F. 1997. Sentinel-Lymph-Node Biopsy. *The New England Journal of Medicine,* 340**,** 317.

SCHILLEN, T. B. 1991. Designing a neural network simulator-the MENS modelling environment for network systems: II. *Cabios,* 7**,** 431-446.

SCHMIDT-KITTLER, O., RAGG, T., DASKALAKIS, A., GRANZOW, M., AHR, A., BLANKENSTEIN, T. J., KAUFMANN, M., DIEBOLD, J., ARNHOLDT, H., MULLER, P., BISCHOFF, J., HARICH, D., SCHLIMOK, G., RIETHMULLER, G., EILS, R. & KLEIN, C. A. 2003. From latent disseminated cells to overt metastasis: genetic analysis of systemic breast cancer progression. *Proc Natl Acad Sci U S A,* 100**,** 7737-42.

SCHRENK, P., RIEGER, R., SHAMIYEH, A. & WAYAND, W. 2000. Morbidity following sentinel lymph node biopsy versus axillary lymph node dissection for patients with breast carcinoma. *Cancer,* 88**,** 608-614.

SCHUMACHER, M., SCHMOOR, C., SAUERBREI, W., SCHAUER, A., UMMENHOFER, L., GATZEMEIER, W. & RAUSCHECKER, H. 1993. The prognostic effect of histological tumor grade in node-negative breast cancer patients. *Breast Cancer Res Treat,* 25**,** 235-45.

SCHWARTZ, G. F., GIULIANO, A. E. & VERSONESI, U. 2002. Proceedings of the consensus conference on the roles of sentinel lymph node biopsy in carcinoma of the breast, April 19-22, Philaldelphia, Pennsylvania. *American Cancer Society,* 94**,** 2542-2551.

SCHWARTZ, G. F. & MELTZER, A. J. 2003. Accuracy of axillary sentinel lymph node biopsy following neoadjuvant (induction) chemotherapy for carcinoma of the breast. *Breast J,* 9**,** 374-9.

SEBASTIAN, F. S., ALEXANDRA, F., MONIKA, S., THOMAS, B.-H., KATALIN, N., MICHAEL, G., REINHARD, H., RAIMUND, J. & PETER, B. 2006. Hypoxia inducible factor-1α correlates with VEGF-C expression and lymphangiogenesis in breast cancer. *Breast Cancer Research and Treatment,* V99**,** 135-141.

SEIDMAN, A., FORNIER, M., ESTEVA, F., TAN, L. K., KAPTAIN, S., BACH, A., PANAGEAS, K., ARROYO, C., VALERO, V., CURRIE, V., GILEWSKI, T., THEODOULOU, M., MOYNAHAN, M., MOASSER, M., SKLARIN, N., DICKLER, M., D'ANDREA, G., CRISTOFANILLI, M., RIVERA, E., HORTOBAGYI, G., NORTON, L. & HUDIS, C. 2001. Weekly trastuzumab and paclitaxel therapy for metastatic

breast cancer with analysis of efficacy by HER2 immunophenotype and gene amplification. *Journal of Clincal Oncology,* 19**,** 2587-2595.

SEKER, H., ODETAYO, M. O., PETROVIC, D. & NAGUIB, R. N. 2003. A fuzzy logic based-method for prognostic decision making in breast and prostate cancers. *IEEE Trans Inf Technol Biomed,* 7**,** 114-22.

SEKER, H., ODETAYO, M. O., PETROVIC, D., NAGUIB, R. N., BARTOLI, C., ALASIO, L., LAKSHMI, M. S. & SHERBET, G. V. 2002. Assessment of nodal involvement and survival analysis in breast cancer patients using image cytometric data: statistical, neural network and fuzzy approaches. *Anticancer Res,* 22**,** 433-8.

SHERIDAN, C., KISHIMOTO, H., FUCHS, R. K., MEHROTRA, S., BHAT-NAKSHATRI, P., TURNER, C. H., GOULET, R., JR., BADVE, S. & NAKSHATRI, H. 2006. CD44+/CD24- breast cancer cells exhibit enhanced invasive properties: an early step necessary for metastasis. *Breast Cancer Res,* 8**,** R59.

SHI, M., LIU, D., DUAN, H., SHEN, B. & GUO, N. 2010. Metastasis-related miRNAs, active players in breast cancer invasion, and metastasis. *Cancer Metastasis Rev,* 29**,** 785-99.

SHIPITSIN, M., CAMPBELL, L. L., ARGANI, P., WEREMOWICZ, S., BLOUSHTAIN-QIMRON, N., YAO, J., NIKOLSKAYA, T., SEREBRYISKAYA, T., BEROUKHIM, R., HU, M., HALUSHKA, M. K., SUKUMAR, S., PARKER, L. M., ANDERSON, K. S., HARRIS, L. N., GARBER, J. E., RICHARDSON, A. L., SCHNITT, S. J., NIKOLSKY, Y., GELMAN, R. S. & POLYAK, K. 2007. Molecular definition of breast tumor heterogeneity. *Cancer Cell,* 11**,** 259-73.

SHIPLEY, J. 2006. Putting the colours into chromogenic in situ hybridization (CISH). *J Pathol,* 210**,** 1-2.

SHULTZ, E. K. 1996. Artificial neural networks: laboratory aid of scorcerer's apprentice. *Clinical Chemistry,* 42**,** 496-497.

SIKAND, K., SLAIBI, J. E., SINGH, R., SLANE, S. D. & SHUKLA, G. C. 2010. miR 488* inhibits androgen receptor expression in prostate carcinoma cells. *Int J Cancer*.

SIMON, R., MIRLACHER, M. & SAUTER, G. 2005. Tissue microarrays. *Methods Mol Med,* 114**,** 257-68.

SINCZAK-KUTA, A., TOMASZEWSKA, R., RUDNICKA-SOSIN, L., OKON, K. & STACHURA, J. 2007. Evaluation of HER2/neu gene amplification in patients with invasive breast carcinoma. Comparison of in situ hybridization methods. *Pol J Pathol,* 58**,** 41-50.

SINGLETARY, S. E., ALLRED, C., ASHLEY, P., BASSETT, L. W., BERRY, D., BLAND, K. I., BORGEN, P. I., CLARK, G. M., EDGE, S. B., HAYES, D. F., HUGHES, L. L., HUTTER, R. V., MORROW, M., PAGE, D. L., RECHT, A., THERIAULT, R. L., THOR, A., WEAVER, D. L., WIEAND, H. S. & GREENE, F. L. 2003. Staging system for breast cancer: revisions for the 6th edition of the AJCC Cancer Staging Manual. *Surg Clin North Am,* 83**,** 803-19.

SINGLETARY, S. E. & GREENE, F. L. 2003. Revision of breast cancer staging: the 6th edition of the TNM Classification. *Semin Surg Oncol,* 21**,** 53-9.

SITTER, B., LUNDGREN, S., BATHEN, T. F., HALGUNSET, J., FJOSNE, H. E. & GRIBBESTAD, I. S. 2006. Comparison of HR MAS MR spectroscopic profiles of breast cancer tissue with clinical parameters. *NMR Biomed,* 19**,** 30-40.

SKOBE, M., HAWIGHORST, T., JACKSON, D. G., PREVO, R., JANES, L., VELASCO, P., RICCARDI, L., ALITALO, K., CLAFFEY, K. & DETMAR, M. 2001. Induction of tumor lymphangiogenesis by VEGF-C promotes breast cancer metastasis. *Nat Med,* 7**,** 192-8.

SLAMON, D., CLARK, G., WONG, S., LEVIN, W., ULLRICH, A. & MCGUIRE, W. 1987. Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science,* 235**,** 177-182.

SLAMON, D., GODOLPHIN, W., JONES, L., HOLT, J., WONG, S., KEITH, D., LEVIN, W., STUART, S., UDOVE, J., ULLRICH, A. & PRESS, M. 1989. Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer. *Science,* 244**,** 707-712.

SLEEMAN, J. P. 2000. The lymph node as a bridgehead in the metastatic dissemination of tumors. *Recent Results Cancer Res,* 157**,** 55-81.

SLEEMAN, J. P., NAZARENKO, I. & THIELE, W. 2011. Do all roads lead to Rome? Routes to metastasis development. *Int J Cancer,* 128**,** 2511-26.

SMIDT, M. L., STROBBE, L. J., GROENEWOUD, H. M., DER WILT, G. J., VAN ZEE, K. J. & WOBBES, T. 2007. Can surgical oncologists reliably predict the likelihood for non-SLN metastases in breast cancer patients? *Ann Surg Oncol,* 14**,** 615-20.

SORLIE, T., PEROU, C. M., TIBSHIRANI, R., AAS, T., GEISLER, S., JOHNSEN, H., HASTIE, T., EISEN, M. B., VAN DE RIJN, M., JEFFREY, S. S., THORSEN, T., QUIST, H., MATESE, J. C., BROWN, P. O., BOTSTEIN, D., EYSTEIN LONNING, P. & BORRESEN-DALE, A. L. 2001. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A,* 98**,** 10869-74.

SOTIRIOU, C., NEO, S. Y., MCSHANE, L. M., KORN, E. L., LONG, P. M., JAZAERI, A., MARTIAT, P., FOX, S. B., HARRIS, A. L. & LIU, E. T. 2003. Breast cancer classification and prognosis based on gene expression profiles from a population-based study. *Proc Natl Acad Sci U S A,* 100**,** 10393-8.

SOTIRIOU, C., WIRAPATI, P., LOI, S., HARRIS, A., FOX, S., SMEDS, J., NORDGREN, H., FARMER, P., PRAZ, V., HAIBE-KAINS, B., DESMEDT, C., LARSIMONT, D., CARDOSO, F., PETERSE, H., NUYTEN, D., BUYSE, M., VAN DE VIJVER, M. J., BERGH, J., PICCART, M. & DELORENZI, M. 2006. Gene Expression Profiling in Breast Cancer: Understanding the Molecular Basis of Histologic Grade To Improve Prognosis. *J Natl Cancer Inst,* 98**,** 262-272.

STEINHOFF, M. M. 1999. Axillary Node Micrometastases: Detection and Biologic Significance. *The Breast Journal,* 5**,** 325-329.

STINGL, J. & CALDAS, C. 2007. Molecular heterogeneity of breast carcinomas and the cancer stem cell hypothesis. *Nat Rev Cancer,* 7**,** 791-9.

STIRZAKER, C., ZOTENKO, E., SONG, J. Z., QU, W., NAIR, S. S., LOCKE, W. J., STONE, A., ARMSTONG, N. J., ROBINSON, M. D., DOBROVIC, A., AVERY-KIEJDA, K. A., PETERS, K. M., FRENCH, J. D., STEIN, S., KORBIE, D. J., TRAU, M., FORBES, J. F., SCOTT, R. J., BROWN, M. A., FRANCIS, G. D. & CLARK, S. J. 2015. Methylome sequencing in triple-negative breast cancer reveals distinct methylation clusters with prognostic value. *Nat Commun,* 6**,** 5899.

SUZUKI, M. & TARIN, D. 2007. Gene expression profiling of human lymph node metastases and matched primary breast carcinomas: Clinical implications. *Molecular Oncology,* 1**,** 172-180.

TAHARA, T., YAMAMOTO, E., MADIREDDI, P., SUZUKI, H., MARUYAMA, R., CHUNG, W., GARRIGA, J., JELINEK, J., YAMANO, H. O., SUGAI, T., KONDO, Y., TOYOTA, M., ISSA, J. P. & ESTECIO, M. R. 2014. Colorectal carcinomas with CpG island methylator phenotype 1 frequently contain mutations in chromatin regulators. *Gastroenterology,* 146**,** 530-38 e5.

TAKANE, K. K. & MCPHAUL, M. J. 1996. Functional analysis of the human androgen receptor promoter. *Mol Cell Endocrinol,* 119**,** 83-93.

TALISUNA, A. O., LANGI, P., MUTABINGWA, T. K., WATKINS, W., VAN MARCK, E., EGWANG, T. G. & D'ALESSANDRO, U. 2003. Population-based validation of dihydrofolate reductase gene mutations for the prediction of sulfadoxine-pyrimethamine resistance in Uganda. *Trans R Soc Trop Med Hyg,* 97**,** 338-42.

TARIN, D. 2006. New insights into the pathogenesis of breast cancer metastasis. *Breast Dis,* 26**,** 13-25.

TAVAZOIE, S. F., ALARCON, C., OSKARSSON, T., PADUA, D., WANG, Q., BOS, P. D., GERALD, W. L. & MASSAGUE, J. 2008. Endogenous human microRNAs that suppress breast cancer metastasis. *Nature,* 451**,** 147-52.

TEIXEIRA, M. R., PANDIS, N. & HEIM, S. 2002. Cytogenetic clues to breast carcinogenesis. *Genes Chromosomes Cancer,* 33**,** 1-16.

TEZ, S., YOLDAS, O., KILIC, Y. A., DIZEN, H. & TEZ, M. 2007. Artificial neural networks for prediction of lymph node status in breast cancer patients. *Medical Hypotheses,* 68**,** 922-923.

THIELE, W. & SLEEMAN, J. P. 2006. Tumor-induced lymphangiogenesis: a target for cancer therapy? *J Biotechnol,* 124**,** 224-41.

THIKE, A. A., YONG-ZHENG CHONG, L., CHEOK, P. Y., LI, H. H., WAI-CHEONG YIP, G., HUAT BAY, B., TSE, G. M., IQBAL, J. & TAN, P. H. 2014. Loss of androgen receptor expression predicts early recurrence in triple-negative and basal-like breast cancer. *Mod Pathol,* 27**,** 352-60.

THOMASSEN, M., TAN, Q., EIRIKSDOTTIR, F., BAK, M., COLD, S. & KRUSE, T. A. 2007. Prediction of metastasis from low-malignant breast cancer by gene expression profiling. *Int J Cancer,* 120**,** 1070-5.

THOR, A., BERRY, D. & BUDMAN, D. 1998. ErbB-2, p53, and eficacy of adjuvant therapy in lymph-node positive breast cancer. *J Natl Cancer Inst,* 90**,** 1346-1360.

THOR, A. D., ENG, C., DEVRIES, S., PATERAKOS, M., WATKIN, W. G., EDGERTON, S., MOORE, D. H., 2ND, ETZELL, J. & WALDMAN, F. M. 2002. Invasive micropapillary carcinoma of the breast is associated with chromosome 8 abnormalities detected by comparative genomic hybridization. *Hum Pathol,* 33**,** 628-31.

THORNTON, H. 1999. Sentinel node biopsy in breast cancer. *BMJ,* 318**,** 599.

TODOROVIC-RAKOVIC, N., JOVANOVIC, D., NESKOVIC-KONSTANTINOVIC, Z. & NIKOLIC-VUKOSAVLJEVIC, D. 2007. Prognostic value of HER2 gene amplification detected by chromogenic in situ hybridization (CISH) in metastatic breast cancer. *Exp Mol Pathol,* 82**,** 262-8.

TORHORST, J., BUCHER, C., KONONEN, J., HAAS, P., ZUBER, M., KOCHLI, O. R., MROSS, F., DIETERICH, H., MOCH, H., MIHATSCH, M., KALLIONIEMI, O. P. & SAUTER, G. 2001. Tissue microarrays for rapid linking of molecular changes to clinical endpoints. *Am J Pathol,* 159**,** 2249-56.

TORLAKOVIC, E. E., FRANCIS, G., GARRATT, J., GILKS, B., HYJEK, E., IBRAHIM, M., MILLER, R., NIELSEN, S., PETCU, E. B., SWANSON, P. E., TAYLOR, C. R., VYBERG, M. & INTERNATIONAL

AD HOC EXPERT, P. 2014. Standardization of negative controls in diagnostic immunohistochemistry: recommendations from the international ad hoc expert panel. *Appl Immunohistochem Mol Morphol,* 22**,** 241-52.

TORLAKOVIC, E. E., NIELSEN, S., FRANCIS, G., GARRATT, J., GILKS, B., GOLDSMITH, J. D., HORNICK, J. L., HYJEK, E., IBRAHIM, M., MILLER, K., PETCU, E., SWANSON, P. E., ZHOU, X., TAYLOR, C. R. & VYBERG, M. 2015. Standardization of positive controls in diagnostic immunohistochemistry: recommendations from the International Ad Hoc Expert Committee. *Appl Immunohistochem Mol Morphol,* 23**,** 1-18.

TORRES, L., RIBEIRO, F. R., PANDIS, N., ANDERSEN, J. A., HEIM, S. & TEIXEIRA, M. R. 2007. Intratumor genomic heterogeneity in breast cancer with clonal divergence between primary carcinomas and lymph node metastases. *Breast Cancer Res Treat,* 102**,** 143-55.

TOUSSAINT, J., SIEUWERTS, A. M., HAIBE-KAINS, B., DESMEDT, C., ROUAS, G., HARRIS, A. L., LARSIMONT, D., PICCART, M., FOEKENS, J. A., DURBECQ, V. & SOTIRIOU, C. 2009. Improvement of the clinical applicability of the Genomic Grade Index through a qRT-PCR test performed on frozen and formalin-fixed paraffin-embedded tissues. *BMC Genomics,* 10**,** 424.

TOYOTA, M., AHUJA, N., OHE-TOYOTA, M., HERMAN, J. G., BAYLIN, S. B. & ISSA, J. P. 1999. CpG island methylator phenotype in colorectal cancer. *Proc Natl Acad Sci U S A,* 96**,** 8681-6.

TRUONG, H., MORIMOTO, R., WALTS, A. E., ERLER, B. & MARCHEVSKY, A. 1995. Neural networks as an aid in the diagnosis of lymphocyte-rich effusions. *Analyt Quant Cytol Histol,* 17**,** 48-54.

TSANG, W. Y. W. 1999. Sentinel lymph node biopsy: not the state of the art in the management of breast cancer yet. *Advances in Anatomic Pathology,* 6**,** 92-96.

TURASHVILI, G., BOUCHAL, J., BAUMFORTH, K., WEI, W., DZIECHCIARKOVA, M., EHRMANN, J., KLEIN, J., FRIDMAN, E., SKARDA, J., SROVNAL, J., HAJDUCH, M., MURRAY, P. & KOLAR, Z. 2007. Novel markers for differentiation of lobular and ductal invasive breast carcinomas by laser microdissection and microarray analysis. *BMC Cancer,* 7**,** 55.

UMETANI, N., GIULIANO, A. E., HIRAMATSU, S. H., AMERSI, F., NAKAGAWA, T., MARTINO, S. & HOON, D. S. 2006. Prediction of breast tumor progression by integrity of free circulating DNA in serum. *J Clin Oncol,* 24**,** 4270-6.

UMETANI, N., MORI, T., KOYANAGI, K., SHINOZAKI, M., KIM, J., GIULIANO, A. E. & HOON, D. S. 2005. Aberrant hypermethylation of ID4 gene promoter region increases risk of lymph node metastasis in T1 breast cancer. *Oncogene,* 24**,** 4721-7.

UNG, O. A. & WETZIG, N. R. 1999. Sentinel node biopsy: evaluating a new technique: Can we safely avoid axillary clearance in selected woman with breast cancer? *MJA,* 171**,** 452-453.

URQUIDI, V. & GOODISON, S. 2007. Genomic signatures of breast cancer metastasis. *Cytogenet Genome Res,* 118**,** 116-29.

VALASTYAN, S., BENAICH, N., CHANG, A., REINHARDT, F. & WEINBERG, R. A. 2009a. Concomitant suppression of three target genes can explain the impact of a microRNA on metastasis. *Genes Dev,* 23**,** 2592-7.

VALASTYAN, S., REINHARDT, F., BENAICH, N., CALOGRIAS, D., SZASZ, A. M., WANG, Z. C., BROCK, J. E., RICHARDSON, A. L. & WEINBERG, R. A. 2009b. A pleiotropically acting microRNA, miR-31, inhibits breast cancer metastasis. *Cell,* 137**,** 1032-46.

VALASTYAN, S. & WEINBERG, R. A. 2010. miR-31: a crucial overseer of tumor metastasis and other emerging roles. *Cell Cycle,* 9**,** 2124-9.

VAN 'T VEER, L. J., DAI, H., VAN DE VIJVER, M. J., HE, Y. D., HART, A. A., MAO, M., PETERSE, H. L., VAN DER KOOY, K., MARTON, M. J., WITTEVEEN, A. T., SCHREIBER, G. J., KERKHOVEN, R. M., ROBERTS, C., LINSLEY, P. S., BERNARDS, R. & FRIEND, S. H. 2002. Gene expression profiling predicts clinical outcome of breast cancer. *Nature,* 415**,** 530-6.

VAN BEERS, E. H. & NEDERLOF, P. M. 2006. Array-CGH and breast cancer. *Breast Cancer Res,* 8**,** 210.

VAN DE VIJVER, M. 2002. Emerging technologies for HER2 testing. *Oncology,* 63(suppl 1)**,** 33-38.

VAN DE VIJVER, M. J., HE, Y. D., VAN'T VEER, L. J., DAI, H., HART, A. A., VOSKUIL, D. W., SCHREIBER, G. J., PETERSE, J. L., ROBERTS, C., MARTON, M. J., PARRISH, M., ATSMA, D., WITTEVEEN, A., GLAS, A., DELAHAYE, L., VAN DER VELDE, T., BARTELINK, H., RODENHUIS, S., RUTGERS, E. T., FRIEND, S. H. & BERNARDS, R. 2002. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med,* 347**,** 1999-2009.

VAN DEN EYNDEN, G. G., VAN DER AUWERA, I., VAN LAERE, S. J., COLPAERT, C. G., VAN DAM, P., DIRIX, L. Y., VERMEULEN, P. B. & VAN MARCK, E. A. 2006. Distinguishing blood and lymph vessel invasion in breast cancer: a prospective immunohistochemical study. *Br J Cancer*.

VAN DER SCHAFT, D. W., PAUWELS, P., HULSMANS, S., ZIMMERMANN, M., VAN DE POLL-FRANSE, L. V. & GRIFFIOEN, A. W. 2007. Absence of lymphangiogenesis in ductal breast cancer at the primary tumor site. *Cancer Lett,* 254**,** 128-36.

VAN DER STEEN, N., PAUWELS, P., GIL-BAZO, I., CASTANON, E., RAEZ, L., CAPPUZZO, F. & ROLFO, C. 2015. cMET in NSCLC: Can We Cut off the Head of the Hydra? From the Pathway to the Resistance. *Cancers (Basel),* 7**,** 556-73.

VAN DEURZEN, C. H., VAN HILLEGERSBERG, R., HOBBELINK, M. G., SELDENRIJK, C. A., KOELEMIJ, R. & VAN DIEST, P. J. 2007. Predictive value of tumor load in breast cancer sentinel lymph nodes for second echelon lymph node metastases. *Cell Oncol,* 29**,** 497-505.

VAN ZEE, K. J., MANASSEH, D. M., BEVILACQUA, J. L., BOOLBOL, S. K., FEY, J. V., TAN, L. K., BORGEN, P. I., CODY, H. S., 3RD & KATTAN, M. W. 2003. A nomogram for predicting the likelihood of additional nodal metastases in breast cancer patients with a positive sentinel node biopsy. *Ann Surg Oncol,* 10**,** 1140-51.

VANHARANTA, S. & MASSAGUE, J. 2013. Origins of metastatic traits. *Cancer Cell,* 24**,** 410-21.

VERONESI, U. 1997. Sentinel-node biopsy in breast cancer - correspondence. *The Lancet,* 350**,** 808-809.

VERONESI, U., MARUBINI, E., MARIANI, L., VALAGUSSA, P. & ZUCALI, R. 1999. The dissection of internal mammary nodes does not improve the survival of breast cancer patients. 30-year results of a randomised trial. *Eur J Cancer,* 35**,** 1320-5.

VERONESI, U., PAGANELLI, G., GALIMBERTI, V., VIALE, G., ZURRIDA, S., BEDONI, M., COSTA, A., CICCO, C. D., GERAGHTY, J. G., LUINI, A., SACCHINI, V. & VERONESI, P. 1997. Sentinel-node biopsy to avoid axillary dissection in breast cancer with clinically negative lymph-nodes. *The Lancet,* 349**,** 1864-67.

VIALE, G. 2012. The current state of breast cancer classification. *Ann Oncol,* 23 Suppl 10**,** x207-10.

VLEUGEL, M. M., BOS, R., VAN DER GROEP, P., GREIJER, A. E., SHVARTS, A., STEL, H. V., VAN DER WALL, E. & VAN DIEST, P. J. 2004. Lack of lymphangiogenesis during breast carcinogenesis. *J Clin Pathol,* 57**,** 746-51.

WADDELL, N., STEIN, S. R., WAGNER, S. A., BENNETT, I., DJOUGARIAN, A., MELANA, S., JAFFER, S., HOLLAND, J. F., POGO, B. G., GONDA, T. J., BROWN, M. A., LEO, P., SAUNDERS, N. A., MCMILLAN, N. A., COCCIARDI, S., VARGAS, A. C., LAKHANI, S. R., CHENEVIX-TRENCH, G., NEWMAN, B. & FRANCIS, G. D. 2012. Morphological and molecular analysis of a breast cancer cluster at the ABC Studio in Toowong. *Pathology,* 44**,** 469-72.

WALTERING, K. K., WALLEN, M. J., TAMMELA, T. L., VESSELLA, R. L. & VISAKORPI, T. 2006. Mutation screening of the androgen receptor promoter and untranslated regions in prostate cancer. *Prostate,* 66**,** 1585-91.

WARD, P. M. & WEISS, L. 1989. Metachronous seeding of lymph node metastases in rats bearing the MT-100-TC mammary carcinoma: the effect of elective lymph node dissection. *Breast Cancer Res Treat,* 14**,** 315-20.

WARFORD, A. 2004. Tissue microarrays: fast-tracking protein expression at the cellular level. *Expert Rev Proteomics,* 1**,** 283-92.

WATANABE, A., CORNELISON, R. & HOSTETTER, G. 2005. Tissue microarrays: applications in genomic research. *Expert Rev Mol Diagn,* 5**,** 171-81.

WEBER-MANGAL, S., SINN, H. P., POPP, S., KLAES, R., EMIG, R., BENTZ, M., MANSMANN, U., BASTERT, G., BARTRAM, C. R. & JAUCH, A. 2003. Breast cancer in young women (< or = 35 years): Genomic aberrations detected by comparative genomic hybridization. *Int J Cancer,* 107**,** 583-92.

WEE, E. J., PETERS, K., NAIR, S. S., HULF, T., STEIN, S., WAGNER, S., BAILEY, P., LEE, S. Y., QU, W. J., BREWSTER, B., FRENCH, J. D., DOBROVIC, A., FRANCIS, G. D., CLARK, S. J. & BROWN, M. A. 2012. Mapping the regulatory sequences controlling 93 breast cancer-associated miRNA genes leads to the identification of two functional promoters of the Hsa-mir-200b cluster, methylation of which is associated with metastasis or hormone receptor status in advanced breast cancer. *Oncogene,* 31**,** 4182-95.

WEIGELT, B., GLAS, A. M., WESSELS, L. F., WITTEVEEN, A. T., PETERSE, J. L. & VAN'T VEER, L. J. 2003. Gene expression profiles of primary breast tumors maintained in distant metastases. *Proc Natl Acad Sci U S A,* 100**,** 15901-5.

WEIGELT, B., HU, Z., HE, X., LIVASY, C., CAREY, L. A., EWEND, M. G., GLAS, A. M., PEROU, C. M. & VAN'T VEER, L. J. 2005a. Molecular portraits and 70-gene prognosis signature are preserved throughout the metastatic process of breast cancer. *Cancer Res,* 65**,** 9155-8.

WEIGELT, B., PETERSE, J. L. & VAN 'T VEER, L. J. 2005b. Breast cancer metastasis: markers and models. *Nat Rev Cancer,* 5**,** 591-602.

WEIGELT, B., WESSELS, L. F., BOSMA, A. J., GLAS, A. M., NUYTEN, D. S., HE, Y. D., DAI, H., PETERSE, J. L. & VAN'T VEER, L. J. 2005c. No common denominator for breast cancer lymph node metastasis. *Br J Cancer,* 93**,** 924-32.

WELFARE, A. I. O. H. A. 2000. Cancer. Canberra: Commonwealth Government Publications.

WERLING, R. W., HWANG, H., YAZIJI, H. & GOWN, A. M. 2003. Immunohistochemical distinction of invasive form nonivasive breast lesions. *Am J Surg Pathol,* 27**,** 82-90.

WEST, M., BLANCHETTE, C., DRESSMAN, H., HUANG, E., ISHIDA, S., SPANG, R., ZUZAN, H., OLSON, J. A., JR., MARKS, J. R. & NEVINS, J. R. 2001. Predicting the clinical status of human breast cancer by using gene expression profiles. *Proc Natl Acad Sci U S A,* 98**,** 11462-7.

WICHA, M. S. 2007. Breast cancer stem cells: the other side of the story. *Stem Cell Rev,* 3**,** 110-2; discussion 113.

WILDING, P., MORGAN, M., GRYGOTIS, A., SHOFFNER, M. & ROSATO, E. 1994. Application of backpropagation neural networks to diagnosis of breast and ovarian cancer. *Cancer,* 77**,** 145-153.

WILLIAMS, C. S., LEEK, R. D., ROBSON, A. M., BANERJI, S., PREVO, R., HARRIS, A. L. & JACKSON, D. G. 2003. Absence of lymphangiogenesis and intratumoural lymph vessels in human metastatic breast cancer. *J Pathol,* 200**,** 195-206.

WOELFLE, U., CLOOS, J., SAUTER, G., RIETHDORF, L., JANICKE, F., VAN DIEST, P., BRAKENHOFF, R. & PANTEL, K. 2003. Molecular signature associated with bone marrow micrometastasis in human breast cancer. *Cancer Res,* 63**,** 5679-84.

WOJDACZ, T. K. & DOBROVIC, A. 2007. Methylation-sensitive high resolution melting (MS-HRM): a new approach for sensitive and high-throughput assessment of methylation. *Nucleic Acids Res,* 35**,** e41.

WOJDACZ, T. K., HANSEN, L. L. & DOBROVIC, A. 2008. A new approach to primer design for the control of PCR bias in methylation studies. *BMC Res Notes,* 1**,** 54.

WOLFF, A. C., HAMMOND, M. E., SCHWARTZ, J. N., HAGERTY, K. L., ALLRED, D. C., COTE, R. J., DOWSETT, M., FITZGIBBONS, P. L., HANNA, W. M., LANGER, A., MCSHANE, L. M., PAIK, S., PEGRAM, M. D., PEREZ, E. A., PRESS, M. F., RHODES, A., STURGEON, C., TAUBE, S. E., TUBBS, R., VANCE, G. H., VAN DE VIJVER, M., WHEELER, T. M. & HAYES, D. F. 2007. American Society of Clinical Oncology/College of American Pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer. *J Clin Oncol,* 25**,** 118-45.

WONG, S. L. 2002. Optimal use of Sentinel Lymph node biopsy versus axillary lymph nose dissection in patients with breast carcinoma. *Cancer,* 95.

WONG, S. Y. & HYNES, R. O. 2006. Lymphatic or hematogenous dissemination: how does a metastatic tumor cell decide? *Cell Cycle,* 5**,** 812-7.

WU, Y., GIGER, M., DOI, K., VYBORNY, C. J., SCHMIDT, R. & METZ, C. 1993. Artificial neural networks in mammography application to decision making in the diagnosis of breast cancer. *Radiology,* 187**,** 81-87.

WUICIK, L., CAVALLI, L. R., CORNELIO, D. A., SCHMID BRAZ, A. T., BARBOSA, M. L., LIMA, R. S., URBAN, C. A., BLEGGI TORRES, L. F., RIBEIRO, E. M. & CAVALLI, I. J. 2007. Chromosome alterations associated with positive and negative lymph node involvement in breast cancer. *Cancer Genet Cytogenet,* 173**,** 114-21.

XI, L., LYONS-WEILER, J., COELLO, M. C., HUANG, X., GOODING, W. E., LUKETICH, J. D. & GODFREY, T. E. 2005. Prediction of lymph node metastasis by analysis of gene expression profiles in primary lung adenocarcinomas. *Clin Cancer Res,* 11**,** 4128-35.

YANG, J., MANI, S. A., DONAHER, J. L., RAMASWAMY, S., ITZYKSON, R. A., COME, C., SAVAGNER, P., GITELMAN, I., RICHARDSON, A. & WEINBERG, R. A. 2004. Twist, a master regulator of morphogenesis, plays an essential role in tumor metastasis. *Cell,* 117**,** 927-39.

YANG, W., KLOS, K., YANG, Y., SMITH, T. L., SHI, D. & YU, D. 2002. ErbB2 overexpression correlates with increased expression of vascular endothelial growth factors A, C, and D in human breast carcinoma. *Cancer,* 94**,** 2855-61.

YAVUZ, S., PAYDAS, S., DISEL, U., ZORLUDEMIR, S. & ERDOGAN, S. 2005. VEGF-C expression in breast cancer: clinical importance. *Adv Ther,* 22**,** 368-80.

YEAP, B. B., WILCE, J. A. & LEEDMAN, P. J. 2004. The androgen receptor mRNA. *Bioessays,* 26**,** 672-82.

YEH, S., HU, Y. C., WANG, P. H., XIE, C., XU, Q., TSAI, M. Y., DONG, Z., WANG, R. S., LEE, T. H. & CHANG, C. 2003. Abnormal mammary gland development and growth retardation in female mice and MCF7 breast cancer cells lacking androgen receptor. *J Exp Med,* 198**,** 1899-908.

YOUNES, M. & LAUCIRICA, R. 1997. Lack of prognostic significance of histological grade in node-negative invasive breast carcinoma. *Clin Cancer Res,* 3**,** 601-4.

ZAFRANI, B., GERBAULT-SEUREAU, M., MOSSERI, V. & DUTRILLAUX, B. 1992. Cytogenetic study of breast cancer: clinicopathologic significance of homogeneously staining regions in 84 patients. *Hum Pathol,* 23**,** 542-7.

ZHANG, D., SALTO-TELLEZ, M., PUTTI, T. C., DO, E. & KOAY, E. S. 2003. Reliability of tissue microarrays in detecting protein expression and gene amplification in breast cancer. *Mod Pathol,* 16**,** 79-84.

ZHANG, J., LI, Y., HAO, X., ZHANG, Q., YANG, K., LI, L., MA, L., WANG, S. & LI, X. 2011. Recent progress in therapeutic and diagnostic applications of lanthanides. *Mini Rev Med Chem,* 11**,** 678-94.

ZHAO, J., WU, R., AU, A., MARQUEZ, A., YU, Y. & SHI, Z. 2002. Determination of HER2 gene amplification by chromogenic in situ hybridization (CISH) in archival breast carcinoma. *Mod Pathol,* 15**,** 657-665.
ZHU, S., WU, H., WU, F., NIE, D., SHENG, S. & MO, Y. Y. 2008. MicroRNA-21 targets tumor suppressor genes in invasion and metastasis. *Cell Res,* 18**,** 350-9.

Chapter 13

Appendices

13.1. Immunohistochemistry protocols

For IHC tissue sections were dewaxed in xylene, rehydrated in graded alcohols followed by antigen retrieval. The slides were then placed on an automated immunostainer, Ventana ES, Ventana BenchMark or Ventana BenchMark XT (Ventana Medical Systems, AZ) and run using the iView detection kit (Ventana Medical Systems, AZ). Counter staining was performed with haemotoxylin.

IHC staining was evaluated semiquantitatively using a combination of intensity and strength of staining. This scoring system is a modified quickscore system (Detre et al., 1995a). This system utilizes a combination of percentage of cells staining and intensity of staining.

13.2. miRNA expression in LVI

13.2.1. Introduction

Baffa et al (Baffa et al., 2009) performed miRNA analysis on 43 matched primary and lymph node metastases of 10 colorectal carcinomas, 10 urothelial carcinomas, 13 breast carcinomas and 10 lung carcinomas using miRNA microarrays. This study identified 32 miRNAs that were differentially expressed between the 43 primary tumours and the related lymph node metastasis. Some organ specific differences were also identified in the expression profile (Table 6).

Table 6: Thirty-two differentially expressed miRNAs in paired primary and metastatic cancers (Baffa et al., 2009).

Unique id	Fold change*	*p* **value**	*t* **value**
hsa-miR-142-5p	3.3	3.50E-06	4.96
hsa-miR-146a	2.3	2.12E-05	4.5
hsa-miR-150	1.7	<1e-07	6.09
hsa-miR-155	1.7	<1e-07	6.53
hsa-miR-146b-5p	1.5	0.00244	3.12
hsa-miR-10b	1.4	0.04552	2.03
hsa-miR-186	1.4	0.03961	2.09
hsa-miR-518a-3p	1.4	0.03061	2.2
hsa-miR-151-3p	1.4	0.01352	2.52

hsa-miR-125b-2	1.3	0.03246	2.17
hsa-miR-520d-3p	1.3	0.01977	2.38
hsa-miR-342-3p	1.3	7.53E-05	4.16
hsa-miR-149	1.2	0.03197	2.18
hsa-miR-671-3p	1.2	0.02785	2.24
hsa-let-7i	1.1	0.001	3.44
hsa-miR-99b	0.9	0.00865	−2.69
hsa-miR-145	0.9	0.04476	−2.04
hsa-miR-30d	0.8	0.00041	−3.67
hsa-miR-30a	0.8	0.00892	−2.68
hsa-miR-24	0.8	0.03506	−2.14
hsa-miR-93	0.8	0.0425	−2.06
hsa-miR-133b	0.8	0.04423	−2.04
hsa-miR-141	0.7	0.00101	−3.41
hsa-miR-130b	0.7	0.00102	−3.4
hsa-miR-133a	0.7	0.0187	−2.4
hsa-miR-143	0.7	0.0227	−2.32
hsa-miR-340	0.6	0.00023	−3.84
hsa-miR-200b	0.6	0.00033	−3.74
hsa-miR-9	0.6	0.00735	−2.75
hsa-miR-30e	0.6	0.00806	−2.71
hsa-miR-203	0.6	0.01638	−2.45
hsa-miR-200c	0.5	0.00027	−3.8

* Presented as actual change in expression.

This paper presented data for the dysregulation of miRNAs, some of these having been reported previously in the literature (Sayed and Abdellatif, 2011) and confirmed the involvement of miRNAs in invasion and metastasis (Gotte, 2010).

13.2.2. Methods

13.6.1.1. miRNA sequencing protocol

13.6.1.1.1. Procedure overview

The procedure is based on Applied Biosystems Ligase-Enhanced Genome Detection (LEGenD™) technology (patent pending).

13.6.1.1.2. Hybridization and ligation to Adaptor Mix

First, the small RNA sample is hybridized with either Adaptor Mix A or Adaptor Mix B. These Adaptor Mixes are sets of oligonucleotides with a single-stranded degenerate sequence at one end and a defined sequence required for SOLiD sequencing at the other end. Each Adaptor Mix constrains the orientation of the

RNA in the ligation reaction such that hybridization with Adaptor Mix A yields template for SOLiD sequencing from the 5' end of the sense strand, while hybridization with Adaptor Mix B yields template for sequencing the reverse
complement (yielding sequence starting from the 3' end of the sense strand).
Note: To achieve higher confidence in the complete sequence of the larger species of small RNAs in a sample, prepare two separate small RNA libraries, using Adaptor Mix A for one reaction and Adaptor Mix B for the other.
After hybridization, the adaptors are ligated to the small RNA molecules using Ligation Enzyme Mix, which is a mixture of an RNA Ligase and other components. Ligation requires an RNA molecule with a 5'-monophosphate and a 3'-hydroxyl end; therefore, most small RNAs can participate in this reaction, and intact mRNA molecules with a 5' cap structure are excluded.

13.6.1.1.3. Reverse transcription and RNase H digestion

Next, the small RNA population with ligated adaptors is reverse transcribed, to generate cDNA. Treatment with RNase H follows, to digest the RNA from RNA/cDNA duplexes and to reduce the concentration of unligated adaptors and adaptor by-products. At this point, reactions contain cDNA copies of the small RNA molecules in the sample.

13.6.1.1.4. cDNA Library amplification

To meet the sample quantity requirements for SOLiD sequencing, and to append the required terminal sequences to each molecule, the cDNA library is amplified using one of the supplied primer sets and 15–18 cycles of PCR. Limiting the cycle number minimizes the synthesis of spurious PCR products and better preserves the small RNA profile of the sample.
Ten sets of PCR primers are included in the kit; they are identical except for a 6 nt “barcode” sequence on the 3' (reverse) primer. The 5' PCR primer is identical in each set; its sequence corresponds to SOLiD emulsion PCR primer 1. The 3' reverse PCR primers are identical except for a 6 nt “barcode” sequence; the SOLiD emulsion PCR primer 2 sequence (P2) is on the 5' side of the barcode, and the adapter sequence is on the 3' side.

13.6.1.1.5. Amplified library cleanup and size selection by PAGE

The final steps in the procedure involve cleanup and size selection of the amplified cDNA library to concentrate samples and remove PCR by-products.
PCR products ~105–150 bp are isolated, corresponding to inserts derived from the small RNA population.

13.6.1.1.6. SOLiD sample preparation and sequencing

The amplified cDNA library generated with the SOLiD Small RNA Expression Kit is ready for attachment to beads at the emulsion PCR step of the SOLiD sample preparation workflow.

13.6.1.1.7. RNA extraction from LCM FFPE material

13.6.1.1.7.1. Paradise Reagent System

The flow chart illustrates the Paradise Reagent System RNA Extraction/Isolation procedure:

(a) Extract RNA from a CapSure HS LCM Cap.

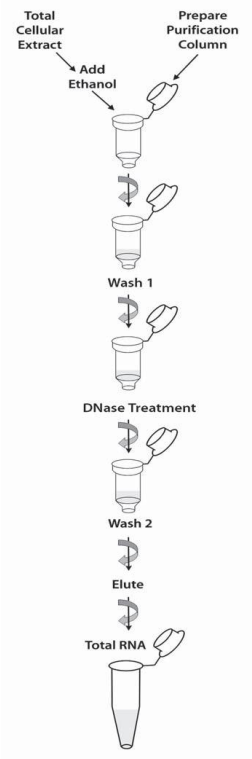
(b) Mix and load cell extract onto a preconditioned purification column.

(c) Spin the extract through the column to capture RNA on the purification column membrane. Wash, DNase treat, and wash again.

(d) Wash the column twice with wash buffer, and

(e) Elute the RNA in low ionic strength buffer.

The entire isolation process, including incubations, can be completed in less than an hour, and the isolated total cellular RNA is ready for use in downstream applications.



Quality control is performed using a tissue scrape protocol. The protocol enables an estimation of RNA quantity and quality using a quantitative real-time PCR assay with primers designed to β-actin. The assumption is that the β-actin mRNA in the sample

represents the average status of other RNA molecules in the same sample. The total estimated RNA amount in a given sample is expressed as an equivalent of universal RNA that contains the same amount of β-actin mRNA.

The protocol measures the average β-actin cDNA length by quantification of the PCR product yield from the 3' end (primer 1650-1717) and another relative 5' sequence (primer 1355-1472). If all cDNA contains both the 3' and 5' sequence target, the ratio of the PCR product for 3'/5' would be 1. As the RNA from FFPE samples tends to exhibit some degradation, the 3'/5' ratio is usually greater than one. Depending on the ratio, an estimation of the quality of the RNA can be made. The reaction mixture is analyzed with an ABI PRISM 7900HT.

For patient samples, cells are laser capture microdissected onto CapSure® LCM Macro caps for primary tumour and lymph nodes. A minimum of 5000 cells were captured using the infrared laser to anneal the cells to the thermoplastic polymer followed by cutting of the membrane by the ultraviolet laser. Cells from lymphatic spaces were identified by dual IHC staining of the initial slide with lymphatic endothelium identified by D2-40 and vascular endothelium by Factor VIII. The sequential H & E, dual IHC and LCM slides are correlated on the Arcturus XT instrument to identify lymphatic spaces. RNA was extracted from a CapSure LCM Cap using proteinase K. The cell extract is loaded onto a preconditioned purification column. The column is centrifuged to capture RNA on the purification column membrane. It is then washed, treated with DNase and the RNA eluted in low ionic strength buffer.

Linear amplification of mRNA was performed using the Arcturus Paradise system followed by gene expression profiling using an Affymetrix platform with X3P chips optimised for LCM from FFPE material.

The Paradise® Plus Reagent System RNA Amplification reagents are optimized to amplify formalin fixed RNA. The reagents utilize two rounds of a five-step process for linear amplification of the mRNA fraction of total cellular RNA:

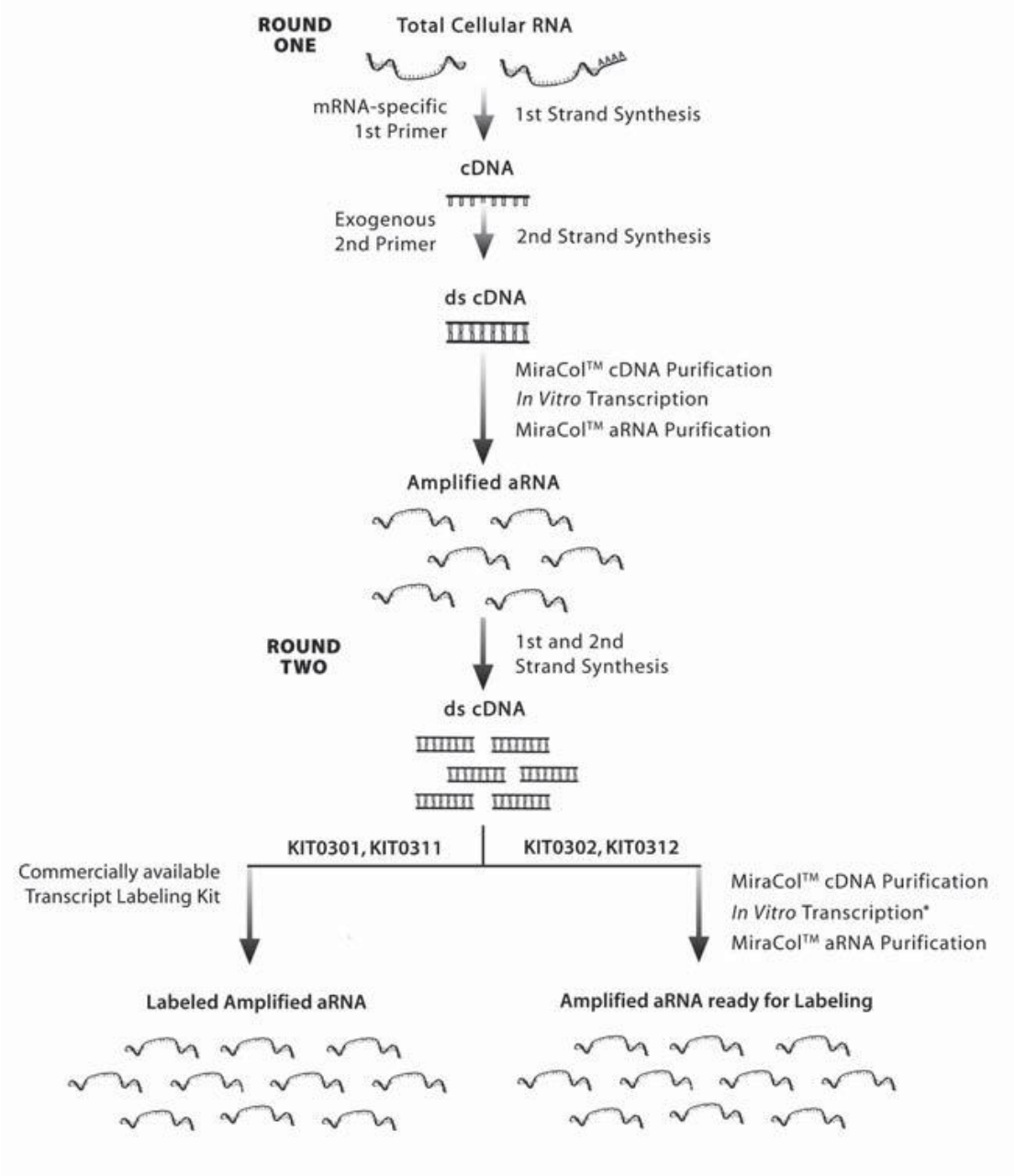
a first-strand synthesis reaction that yields cDNA incorporating a T7 promoter sequence;

b second-strand synthesis reaction utilizing exogenous primers that yields doublestranded cDNA;

c cDNA purification using specially designed MiraCol™ Purification Columns;

d in vitro transcription (IVT) utilizing T7 RNA polymerase yields antisense RNA (aRNA); and

e aRNA isolation with the MiraCol Purification Columns.



Labelling of the product was performed with a number of different protocols to optimise recovery. These included Arcturus Turbo labelling kit, Enzo Bioarray HighYield RNA

transcript Biotin Labelling kit, Mirus LabelIT Array Biotin Labelling kit and the Genisphere FlashTag Biotin RNA Labelling kit.

13.2.3. Results

miRNA sequencing was performed from matched samples from a single patient: frozen primary breast cancer, laser capture microdissection of FFPE primary tumour, laser capture microdissection of FFPE from tumour cells present within lymphatic channels and laser capture microdissection of FFPE from a lymph node metastasis.

Preliminary analysis was performed on the data (Table 7).

Table 7: Summary of total tags mapped to miRNA.

Total tags mapped to miRNAs				
	LNlcm	**LVIlcm**	**PRIMARYfz**	**PRIMARYlcm**
Mature miRNAs	1,044,224	296,898	46,665	384,727
Pre-miRNAs	1,309,491	333,142	72,412	473,176

Counts were normalised by calculating the reads per 100,000 mapped miRNA reads and the results are shown in Tables 8 and 9.

Table 8: Mature-miRNA reads normalised with comparative data for the four samples sequenced.

miRNA	LN LCM	LVI LCM	Primary Frz	Primary LCM
hsa-miR-21	23,128	140	4,532	9,529
hsa-miR-24-1*	6,086	23,329	328	24,357
hsa-miR-425	7,350	46	3,433	4,620
hsa-miR-370	51	23,450	11	2,221
hsa-miR-29c	4,032	17	6,013	3,872
hsa-miR-30b	2,410	1,478	9,388	4,846
hsa-miR-26b	3,646	4	261	1,980
hsa-miR-26a	3,701	426	660	538
hsa-miR-23b	3,200	1,135	1,125	640
hsa-miR-301a	2,940	12	4,363	1,761
hsa-miR-15b	3,083	10	47	1,808
hsa-miR-23a	2,758	15	5,014	2,073
hsa-miR-7-1*	19	8,190	24	3,551
hsa-miR-130b*	2	6,355	34	4,381
hsa-miR-191	2,749	733	58	890
hsa-miR-10b*	0	5,226	15	4,450
hsa-miR-424	2,565	17	6,264	524
hsa-miR-143	451	3,089	5,582	3,567
hsa-miR-30d	2,453	66	84	383

hsa-let-7g	2,244	8	24	847
hsa-miR-505	215	6,389	9	1,430
hsa-miR-342-3p	2,121	963	255	174
hsa-let-7a*	27	5,793	21	2,153
hsa-miR-126*	1,778	12	4,693	1,284
hsa-miR-205	1,789	158	54	6
hsa-miR-19b	1,578	10	3,255	243
hsa-miR-200a	1,459	4	1,254	134
hsa-miR-4284	1,190	132	26	800
hsa-miR-149	1,036	940	13	406
hsa-miR-30a	1,066	44	3,911	36
hsa-miR-125a-5p	732	482	11	734
hsa-miR-197	963	11	13	3
hsa-miR-660	901	4	4	2
hsa-let-7f	300	595	527	876
hsa-miR-101	376	2	2,072	851
hsa-miR-214	2	1,889	4	653
hsa-miR-451	525	2	5,574	1
hsa-miR-106b	256	1	3,523	831
hsa-miR-151-5p	422	3	6	703
hsa-miR-223	469	32	4	508
hsa-miR-10b	0	2,320	0	1
hsa-miR-182	629	2	4	13
hsa-miR-17	613	3	6	2
hsa-miR-125b	479	57	2	221
hsa-miR-181b	519	0	167	28
hsa-miR-484	423	2	2,289	1
hsa-miR-224	0	628	2	944
hsa-miR-150	511	2	0	2
hsa-miR-29a	337	1	1,024	236
hsa-let-7b	299	0	2,683	55
hsa-miR-345	195	465	0	261
hsa-miR-31*	140	45	6	688
hsa-miR-19a	331	1	13	200
hsa-miR-425*	390	2	0	3
hsa-miR-769-5p	381	2	2	1
hsa-miR-193a-3p	340	32	0	1
hsa-miR-361-5p	284	2	2	174
hsa-miR-92a	296	1	2	135
hsa-miR-186	275	31	9	163
hsa-let-7a	331	0	4	31
hsa-miR-20a	279	0	1,296	0
hsa-miR-15b*	305	0	409	0
hsa-miR-148b	185	1	4	337
hsa-miR-96	236	1	2	189
hsa-miR-145	32	414	3,409	0
hsa-miR-192	0	1	4	812
hsa-miR-25	160	2	2	322
hsa-miR-142-5p	1	0	5,604	0
hsa-miR-487b	0	1	2	654

hsa-miR-93	194	1	171	46
hsa-miR-532-5p	139	208	2	3
hsa-miR-30e	85	1	11	293
hsa-miR-199b-3p	1	1	1,382	324
hsa-miR-200b	3	7	0	475
hsa-miR-24	63	19	6	286
hsa-miR-361-3p	166	0	4	1
hsa-miR-485-3p	0	577	0	1
hsa-miR-126	5	0	3,227	4
hsa-miR-140-5p	0	494	0	0
hsa-miR-221	1	243	1,069	56
hsa-miR-374b	130	0	0	14
hsa-miR-183	0	0	2	361
hsa-miR-217	4	394	2	25
hsa-miR-21*	22	241	17	80
hsa-miR-374b*	0	1	2	325
hsa-miR-190	104	1	223	0
hsa-miR-145*	0	2	4	307
hsa-miR-409-3p	0	188	0	156
hsa-miR-210	0	383	0	2
hsa-miR-18a*	63	0	0	124
hsa-miR-140-3p	0	351	4	0
hsa-miR-100	0	341	0	0
hsa-miR-151-3p	0	1	2	256
hsa-miR-340*	0	4	0	249
hsa-miR-331-3p	10	1	0	224
hsa-let-7i	84	1	51	10
hsa-miR-141	84	16	0	0
hsa-miR-10a	89	0	2	0
hsa-miR-542-5p	0	1	0	238
hsa-miR-27a	0	176	0	88
hsa-miR-148a	0	0	0	210
hsa-miR-363	0	0	0	206
hsa-let-7i*	0	266	0	0
hsa-miR-1248	64	0	0	25
hsa-miR-340	59	0	0	36
hsa-miR-200c	0	0	0	176
hsa-miR-494	0	0	0	172
hsa-miR-33b	61	0	0	0
hsa-miR-199b-5p	0	1	1,301	1
hsa-miR-574-3p	0	0	0	155
hsa-miR-152	0	0	1,219	0
hsa-miR-27b	13	0	0	113
hsa-miR-493	0	191	0	0
hsa-miR-18a	46	0	0	13
hsa-miR-107	44	21	0	3
hsa-miR-181a-2*	0	87	0	64
hsa-let-7d*	0	0	2	128
hsa-miR-203	45	0	0	0
hsa-miR-449a	45	0	0	0

hsa-miR-335	45	0	0	0
hsa-miR-767-5p	0	157	0	0
hsa-miR-15a	0	1	0	117
hsa-miR-450a	0	7	0	102
hsa-miR-148b*	0	24	0	89
hsa-miR-2114	0	0	0	107
hsa-miR-376c	38	0	0	0
hsa-miR-130b	37	0	0	0
hsa-let-7d	0	0	821	0
hsa-miR-497	32	0	0	1
hsa-miR-16-2*	31	0	0	0
hsa-miR-449c	29	1	0	0
hsa-miR-545	0	0	654	0
hsa-miR-590-3p	28	0	0	0
hsa-miR-9*	27	0	0	0
hsa-miR-381	1	74	0	0
hsa-miR-181a*	0	0	0	51
hsa-miR-1287	16	0	0	0
hsa-miR-190b	15	0	6	0
hsa-miR-3607-5p	0	15	221	0
hsa-miR-423-3p	10	12	0	0
hsa-miR-122	0	42	0	3
hsa-miR-99b	0	0	0	31
hsa-miR-34a	8	0	0	5
hsa-miR-142-3p	0	0	0	28
hsa-miR-29c*	0	0	0	25
hsa-miR-193a-5p	0	31	0	1
hsa-miR-504	0	0	0	22
hsa-miR-99b*	0	27	0	0
hsa-miR-375	7	0	0	0
hsa-miR-29b-2*	0	24	0	0
hsa-miR-139-3p	0	23	0	0
hsa-miR-545*	0	0	131	0
hsa-miR-195	0	19	0	0
hsa-miR-1226	0	19	0	0
hsa-miR-421	0	18	0	0
hsa-miR-150*	0	15	0	1
hsa-miR-339-3p	0	10	2	0
hsa-miR-219-1-3p	0	0	0	8
hsa-miR-222	0	0	0	8
hsa-miR-34c-5p	0	3	0	5
hsa-miR-20b	0	7	0	0
hsa-miR-20a*	0	3	0	3
hsa-miR-17*	0	5	0	0
hsa-miR-325	0	3	0	0
hsa-miR-654-3p	1	0	0	0
hsa-miR-22	0	0	0	2
hsa-miR-30d*	1	0	0	0
hsa-miR-374a	0	0	0	2
hsa-miR-671-5p	0	2	0	0

hsa-let-7f-2*	0	2	0	0
hsa-miR-196a*	0	2	0	0
hsa-miR-22*	0	0	11	0
hsa-miR-483-3p	0	0	0	1
hsa-let-7c	0	0	0	0
hsa-miR-34c-3p	0	1	0	0
hsa-miR-29b	0	0	0	0
hsa-miR-542-3p	0	0	0	0
hsa-miR-218	0	0	9	0
hsa-miR-4301	0	0	0	0
hsa-miR-365	0	0	0	0
hsa-miR-624	0	0	0	0
hsa-miR-551b	0	0	0	1
hsa-miR-940	0	2	0	0
hsa-miR-378	0	0	0	0
hsa-miR-4286	0	0	0	0
hsa-miR-1260b	0	1	0	0

Table 9: pre-miRNA reads normalised with comparative data for the four samples sequenced.

miRNA	LN LCM	LVI LCM	Primary Frz	Primary LCM
hsa-mir-21	18,565	343	2,933	7,813
hsa-mir-24-1	4,853	20,791	211	19,804
hsa-mir-29b-1	12,627	135	22,049	7,797
hsa-mir-425	6,172	58	2,214	3,759
hsa-mir-370	41	20,899	7	1,806
hsa-mir-29c	3,215	15	3,876	3,168
hsa-mir-30b	1,922	1,317	6,050	3,940
hsa-mir-181a-2	2,221	1,198	5,234	2,857
hsa-mir-26b	2,907	4	168	1,610
hsa-mir-15b	2,702	9	294	1,470
hsa-mir-30c-2	2,609	64	3,997	1,086
hsa-mir-26a-2	2,951	379	425	437
hsa-mir-23b	2,552	1,011	725	521
hsa-mir-301a	2,345	11	2,813	1,432
hsa-mir-10b	0	6,724	12	3,619
hsa-mir-23a	2,199	14	3,232	1,685
hsa-mir-7-1	15	7,299	15	2,887
hsa-mir-130b	31	5,663	22	3,563
hsa-mir-191	2,192	653	37	723
hsa-mir-424	2,046	15	4,037	426
hsa-mir-143	360	2,753	3,597	2,900
hsa-mir-103-2	1,562	324	2,861	1,308
hsa-let-7a-3	285	5,163	17	1,775
hsa-let-7c	1	4,988	22	2,433
hsa-mir-30d	1,956	59	54	312
hsa-mir-126	1,422	11	5,104	1,048
hsa-let-7g	1,789	7	15	689
hsa-mir-505	172	5,694	6	1,162
hsa-mir-342	1,692	858	164	142

hsa-mir-376b	1	3,073	10	2,304
hsa-mir-205	1,427	141	35	5
hsa-mir-19b-2	1,259	9	2,098	198
hsa-mir-200a	1,163	4	808	109
hsa-mir-4284	949	118	17	650
hsa-mir-149	826	837	8	330
hsa-mir-199a-1	772	4	1,329	478
hsa-mir-30a	850	39	2,520	30
hsa-mir-125a	584	429	7	596
hsa-mir-197	768	10	8	2
hsa-mir-660	719	3	3	1
hsa-let-7f-2	239	532	340	712
hsa-mir-101-2	300	2	1,335	692
hsa-mir-214	1	1,683	3	531
hsa-mir-151	337	3	6	780
hsa-mir-451	419	2	3,592	1
hsa-mir-106b	204	1	2,270	676
hsa-mir-223	374	28	3	413
hsa-mir-182	501	2	3	11
hsa-mir-17	489	7	4	2
hsa-mir-125b-2	382	51	1	180
hsa-mir-181b-2	414	0	108	23
hsa-mir-484	338	2	1,475	1
hsa-mir-224	0	560	1	767
hsa-mir-150	407	15	0	2
hsa-mir-361	358	2	4	142
hsa-mir-29a	268	1	660	192
hsa-let-7b	239	0	1,729	45
hsa-mir-345	156	414	0	212
hsa-mir-145	25	370	2,200	250
hsa-mir-31	112	40	4	559
hsa-mir-19a	264	1	8	163
hsa-mir-769	304	2	1	0
hsa-let-7a-2	287	0	6	25
hsa-mir-193a	271	56	0	1
hsa-mir-148b	148	22	3	346
hsa-mir-92a-2	236	1	1	110
hsa-mir-186	220	28	6	133
hsa-mir-20a	222	3	835	2
hsa-mir-96	188	1	1	154
hsa-mir-192	0	1	3	660
hsa-mir-25	128	2	1	262
hsa-mir-142	0	0	3,611	23
hsa-mir-374b	103	1	1	276
hsa-mir-487b	0	1	1	532
hsa-mir-199b	1	1	1,729	264
hsa-mir-140	0	753	3	0
hsa-mir-421	0	751	0	0
hsa-mir-93	155	2	110	37
hsa-mir-532	111	185	1	2

hsa-mir-30e	68	1	7	238
hsa-mir-200b	2	6	0	386
hsa-mir-24-2	50	17	4	232
hsa-let-7i	67	237	33	8
hsa-mir-340	47	4	0	232
hsa-mir-485	0	514	0	1
hsa-mir-18a	87	0	0	112
hsa-mir-221	1	216	689	46
hsa-mir-183	0	0	1	293
hsa-mir-217	3	351	1	20
hsa-mir-190	83	1	144	0
hsa-mir-409	0	167	0	127
hsa-mir-210	0	341	0	2
hsa-mir-1274b	0	5	0	216
hsa-mir-100	0	304	0	0
hsa-mir-331	8	1	0	182
hsa-mir-141	67	14	0	0
hsa-mir-10a	71	0	1	0
hsa-mir-542	0	1	0	193
hsa-let-7d	0	0	530	104
hsa-mir-27a	0	157	0	72
hsa-mir-148a	0	0	0	171
hsa-mir-363	0	0	0	168
hsa-mir-1248	52	0	0	22
hsa-mir-181b-1	0	0	0	149
hsa-mir-382	0	206	0	0
hsa-mir-200c	0	0	0	143
hsa-mir-494	0	0	0	140
hsa-mir-33b	49	0	0	0
hsa-mir-574	0	0	0	126
hsa-mir-152	0	0	786	0
hsa-mir-493	0	170	0	0
hsa-mir-27b	10	0	0	92
hsa-mir-1274a	0	161	0	0
hsa-mir-107	35	18	0	2
hsa-mir-3607	0	14	146	75
hsa-mir-203	36	1	0	1
hsa-mir-16-1	35	1	11	3
hsa-mir-449a	36	0	0	0
hsa-mir-335	36	0	0	0
hsa-mir-767	0	140	0	0
hsa-mir-15a	0	1	0	95
hsa-mir-450a-2	0	7	0	83
hsa-mir-2114	0	0	0	87
hsa-mir-376c	30	0	0	0
hsa-mir-545	0	0	505	0
hsa-mir-497	26	0	0	0
hsa-mir-16-2	25	0	0	0
hsa-mir-9-2	24	0	3	0
hsa-mir-449c	23	1	0	0

hsa-mir-590	23	0	0	0
hsa-mir-9-3	22	0	0	0
hsa-mir-381	1	66	0	0
hsa-mir-99b	0	24	0	25
hsa-mir-181a-1	0	0	0	41
hsa-mir-1287	13	0	0	0
hsa-mir-190b	12	0	4	0
hsa-mir-423	8	11	0	0
hsa-mir-122	0	37	0	2
hsa-mir-3929	6	3	18	4
hsa-mir-34a	7	0	0	4
hsa-mir-504	0	0	0	18
hsa-mir-3676	3	8	0	1
hsa-mir-375	6	0	0	0
hsa-mir-29b-2	0	22	0	0
hsa-mir-139	0	21	0	0
hsa-mir-1234	0	13	0	3
hsa-mir-195	0	17	0	0
hsa-mir-1226	0	17	0	0
hsa-mir-339	0	9	1	0
hsa-mir-34c	0	3	0	4
hsa-mir-219-1	0	0	0	6
hsa-mir-222	0	0	0	6
hsa-mir-196a-1	2	0	0	0
hsa-mir-3195	0	5	0	1
hsa-mir-20b	0	7	0	0
hsa-mir-620	0	6	0	0
hsa-mir-3648	0	5	0	0
hsa-mir-22	0	0	7	1
hsa-mir-654	1	0	0	0
hsa-mir-325	0	2	0	0
hsa-mir-671	0	2	0	0
hsa-mir-374a	0	0	0	1
hsa-mir-558	0	0	1	0
hsa-mir-196a-2	0	2	0	0
hsa-mir-483	0	0	0	1
hsa-mir-218-2	0	0	6	0
hsa-mir-3908	0	0	1	0
hsa-mir-579	0	0	0	0
hsa-mir-4301	0	0	0	0
hsa-mir-551b	0	0	0	0
hsa-mir-940	0	1	0	0
hsa-mir-3138	0	0	0	0
hsa-mir-4286	0	0	0	0
hsa-mir-1303	0	1	0	0
hsa-mir-378	0	0	0	0
hsa-mir-624	0	0	0	0
hsa-mir-1260b	0	1	0	0
hsa-mir-365-2	0	0	0	0
hsa-mir-1322	0	0	0	0

hsa-mir-198	0	0	0	0
hsa-mir-552	0	0	0	0
hsa-mir-3118-5	0	0	0	0
hsa-mir-1243	0	0	0	0
hsa-mir-1247	0	0	0	0
hsa-mir-663	0	0	0	0
hsa-mir-758	0	0	0	0
hsa-mir-578	0	0	0	0
hsa-mir-92b	0	0	0	0
hsa-mir-153-1	0	0	1	0
hsa-mir-550b-2	0	0	0	0
hsa-mir-628	0	0	0	0
hsa-mir-4268	0	0	0	0
hsa-mir-144	0	0	0	0
hsa-mir-1285-1	0	0	0	0
hsa-mir-580	0	0	0	0
hsa-mir-454	0	0	0	0
hsa-mir-378c	0	0	0	0
hsa-mir-3936	0	0	0	0
hsa-mir-3658	0	0	0	0
hsa-mir-3652	0	0	0	0
hsa-mir-3154	0	0	1	0

A scatterplot of the results shows differences in expression between the different sample types and in particular there are differences in the expression profile for the LVI derived cells compared to both the primary and metastatic tumour (Figure 10).

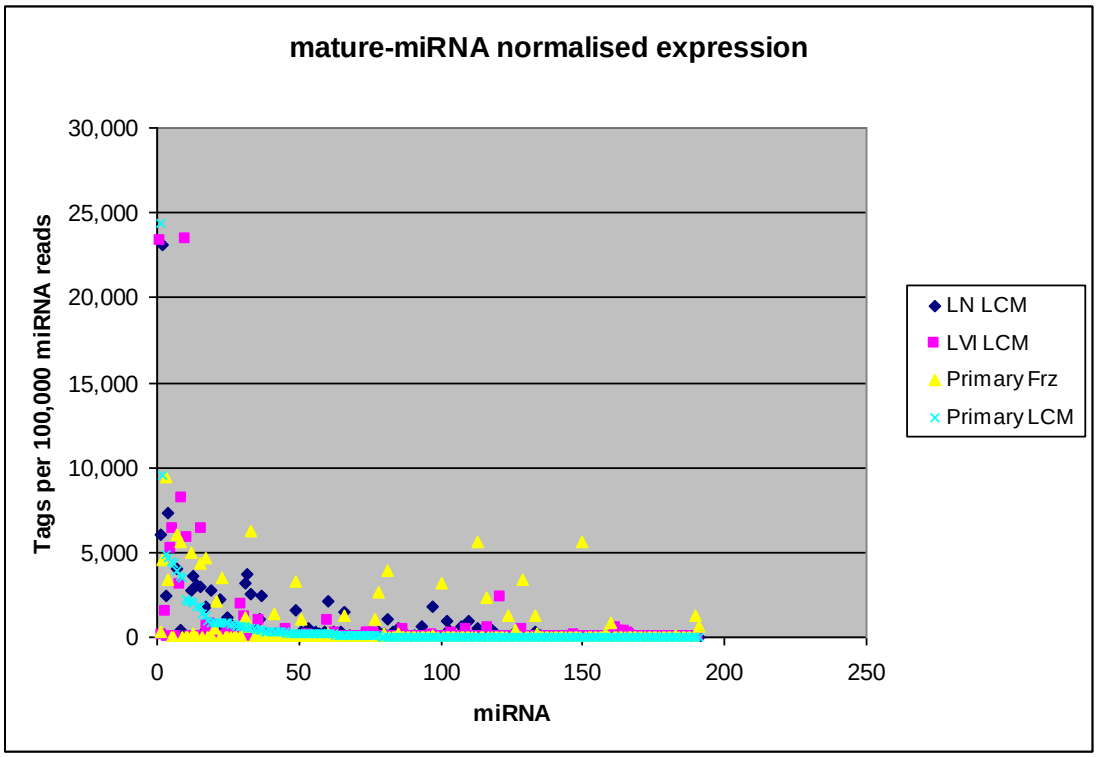


Figure 10: Scatterplot showing normalised expression of miRNAs in different sample types.

There is a reduction in expression of miR-21, miR15b, miR-23a, miR-29c, miR-301a and miR-425 in the LVI cells compared to primary FFPE LCM and lymph node LCM specimens with increased relative expression of miR-10b and miR 370. Expression of miR-21 was higher in the lymph node metastasis compared to the primary tumour.

13.2.4. Discussion

There was a low number of miRNA reads (<1% of the total read count). This was not unexpected as the material was of low yield, being predominantly from LCM FFPE material. LCM collects cells onto a thermoplastic polymer cap for analysis. The number of cells varies in any capture, but is usually only in the 100s-1000s. A mammalian cell contains approximately 10-30 pg of total RNA with mRNA comprising approximately 1-3% of total RNA. miRNA constitutes a small percentage of available material. The majority of RNA in cells is composed of rRNA (more than 80%). Whilst there are methods to remove or reduce the amount of rRNA from samples, these methods rely on intact rRNA to be effective. The crosslinking and fragmentation of nucleic acids in FFPE material makes these approaches ineffective. Therefore the majority of the material sequenced from FFPE material is either fragmented rRNA or empty adapters due to the low amount of starting material. This results in a small percentage of reads mapping back to a reference miRNA database.

This preliminary data from a single patient indicates that it is possible to use next generation sequencing to identify miRNA from LCM FFPE material with low amounts of starting material. The expression data is also consistent with published data in the literature. miR-10b is increased in the LVI cells and has previously been shown to enhance migration and invasion in vitro and metastasis in vivo (Ma et al., 2007a). miR-21 shows a reduced expression in the LVI cells, compared to the primary tumour and there is increased expression in the lymph node metastasis. miR-21 expression is associated with invasion and migration (Hurst et al., 2009b) and high levels of miR-21 expression in breast cancer has been associated with aggressive disease, high tumour grade and negative hormone receptor status (Qian et al., 2009). Data also suggests that miR-21 may have a potential role in promoting tumour invasion and metastasis by simultaneously down-regulating multiple metastasis-related tumour suppressor genes (Zhu et al., 2008).

Glenn Francis SN: 30358409 Protein expression and molecular profiling to predict lymph node status in breast cancer.

13.3. Additional results for breast cancer cluster at the ABC Studio in Toowong.

13.3.1. Results

Images of the TMA sections are illustrated in Figures 11, 12 and 13 for different patient cases.

Figure 11: IHC/ISH stained TMA sections for cases A10595 (left) and A10599 (Right).

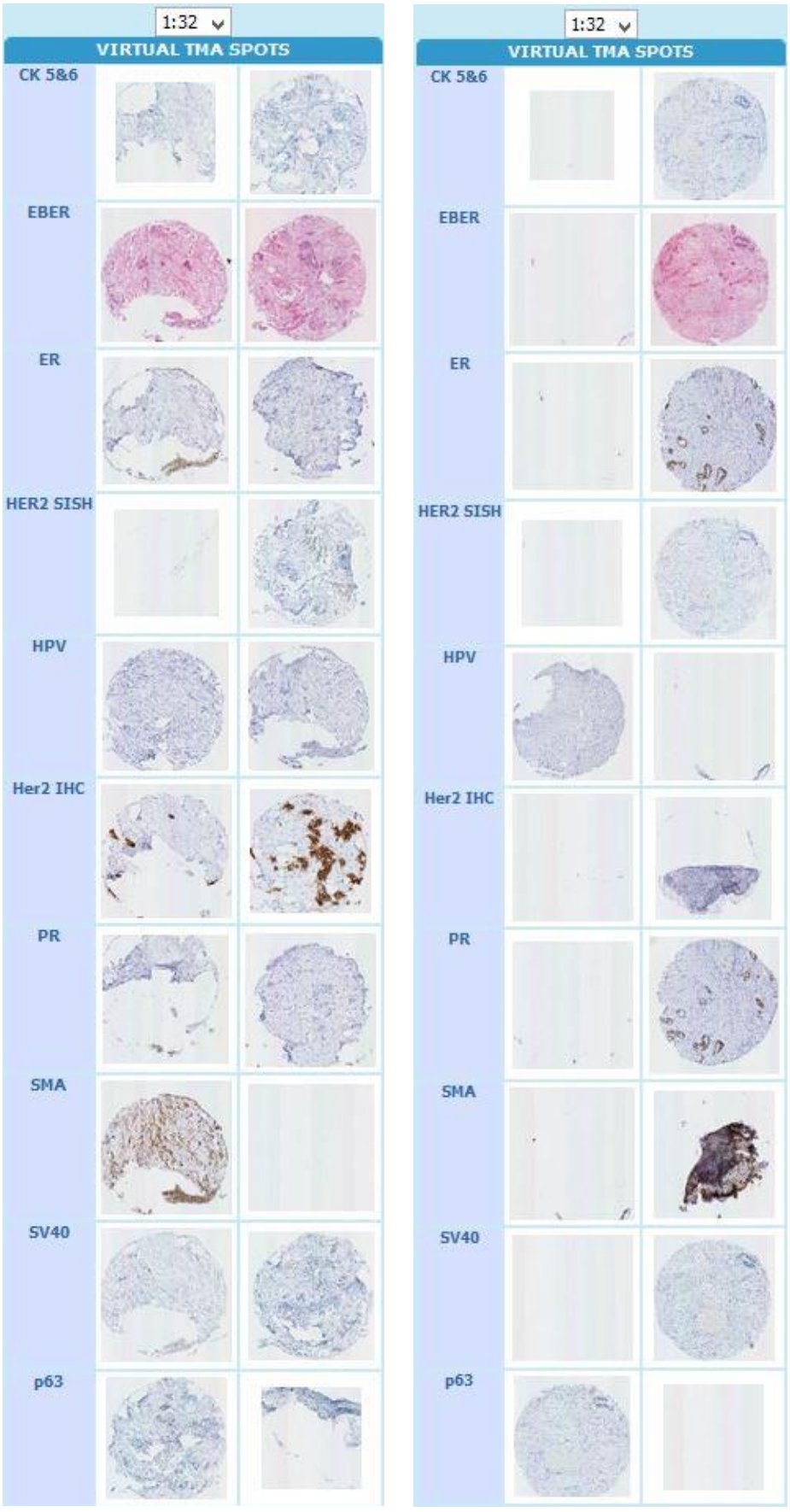


Figure 12: IHC/ISH stained TMA sections for cases A10605 (left) and A10608 (Right).

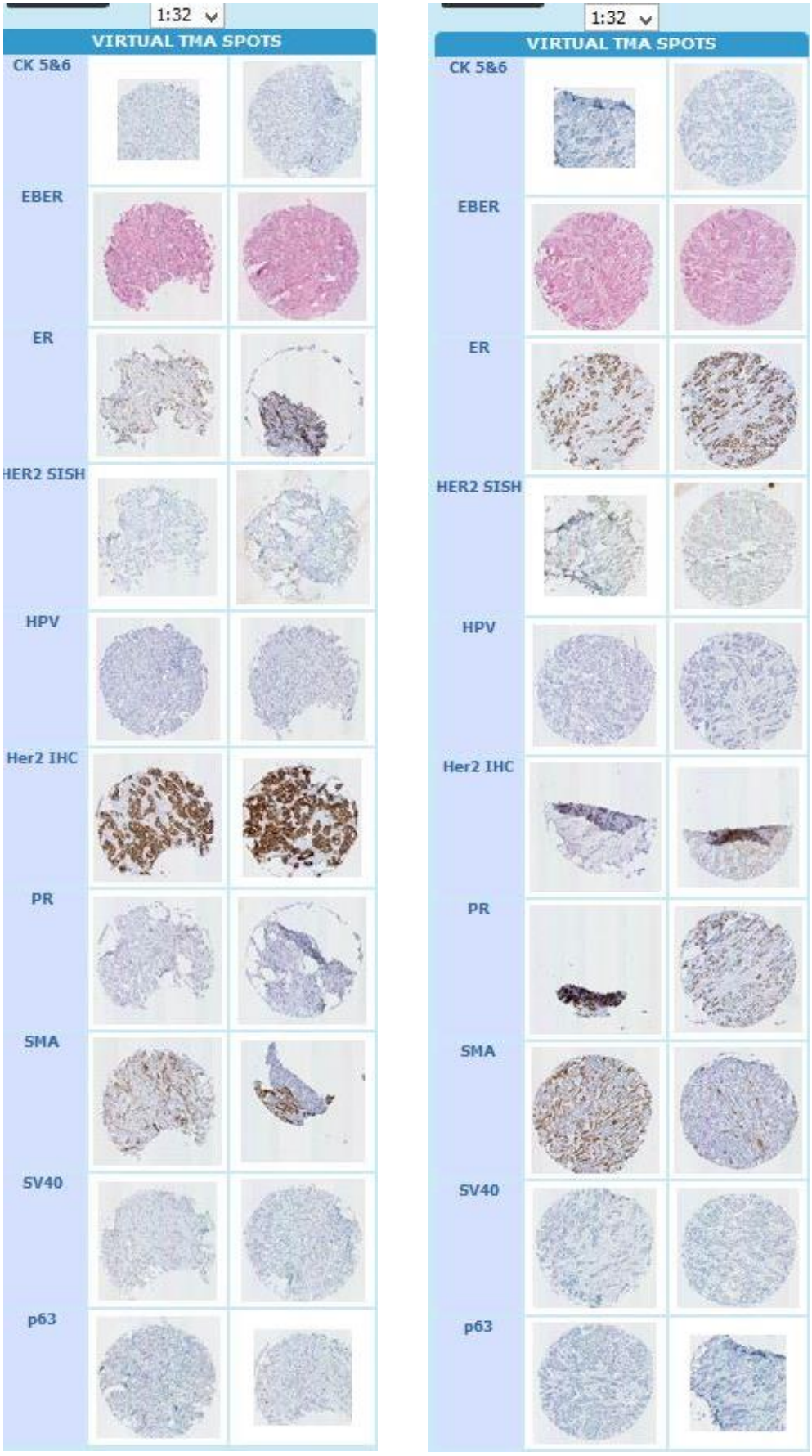
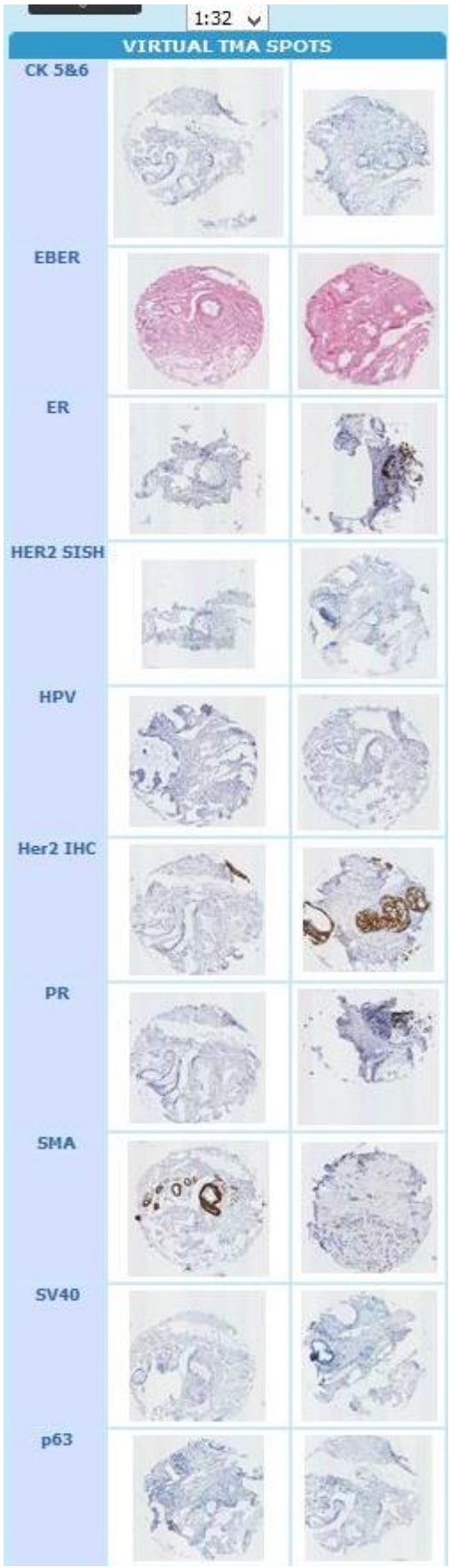


Figure 13: IHC/ISH stained TMA sections for case A10611.



13.4. Additional results for Prediction of Molecular Grade using an Artificial Neural Network (ANN).

13.4.1. Results

TMA results are illustrated for a Grade 3 breast carcinoma case showing IHC for multiple markers including BUB1B, RaCGAP1, RRM2, NEK2 and CENPA (Figures 14, 15 and 16).

Figure 14: IHC on TMAs for multiple markers in a Grade 3 breast carcinoma.

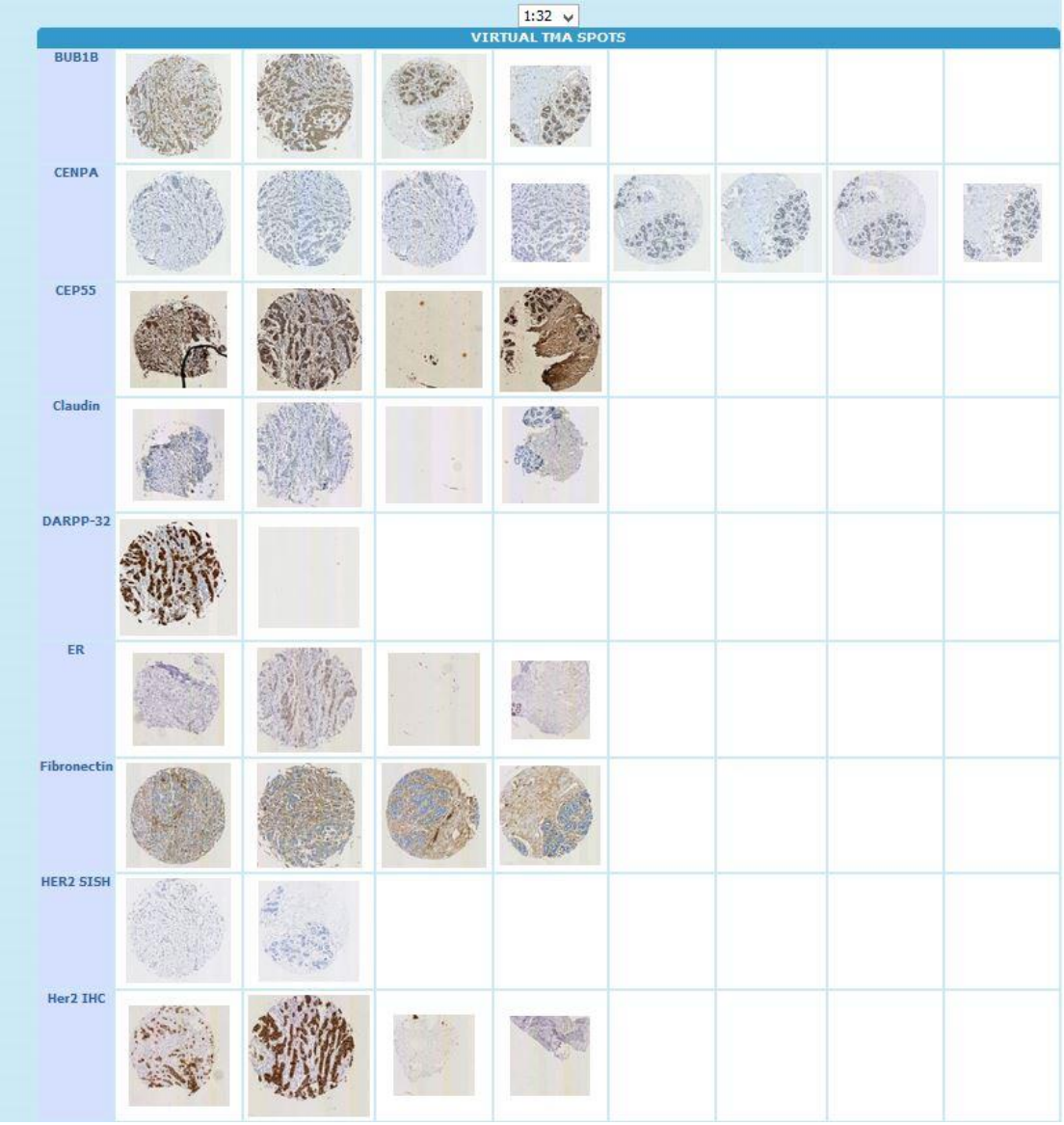


Figure 15: IHC on TMAs for multiple markers in a Grade 3 breast carcinoma.

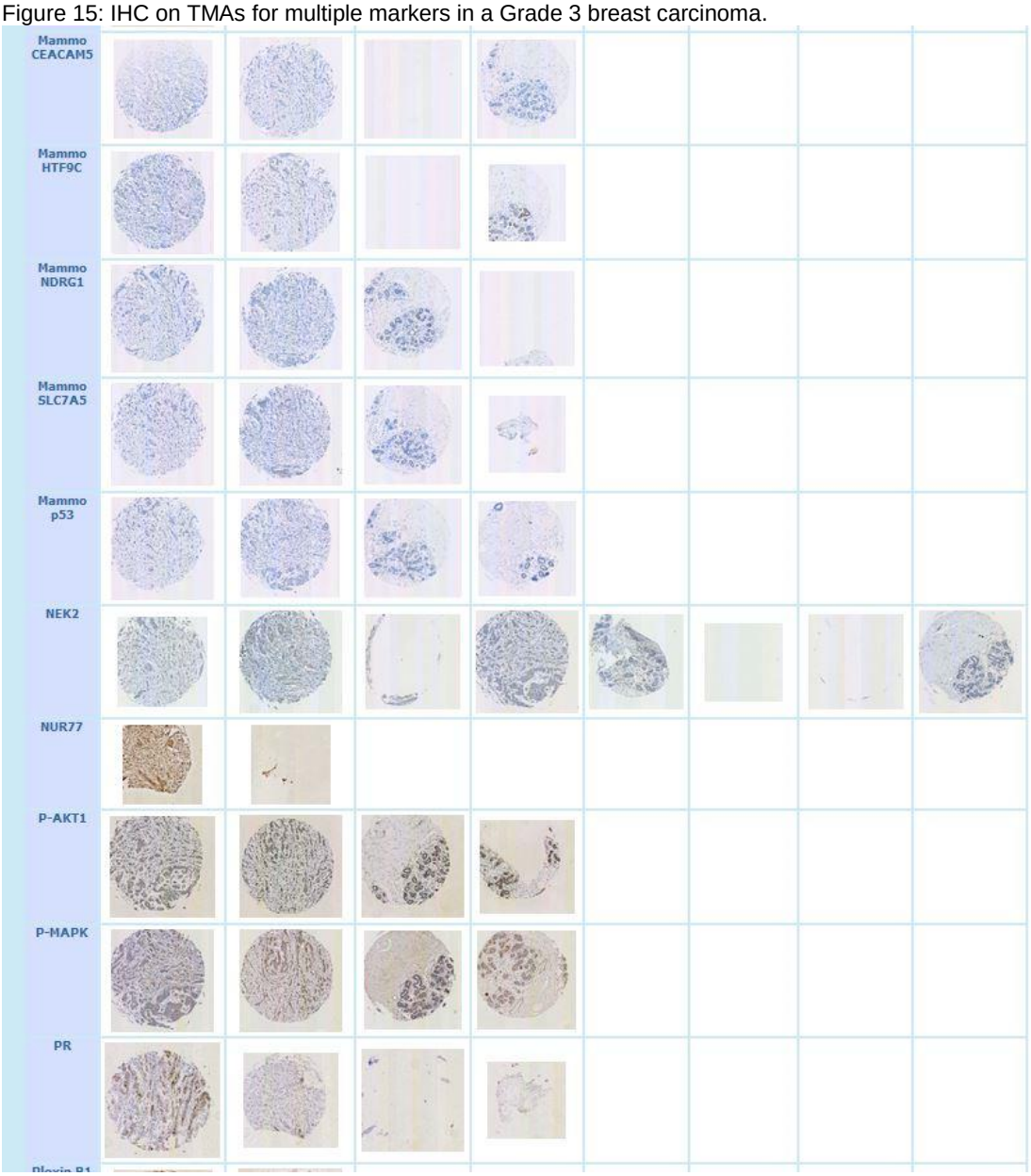
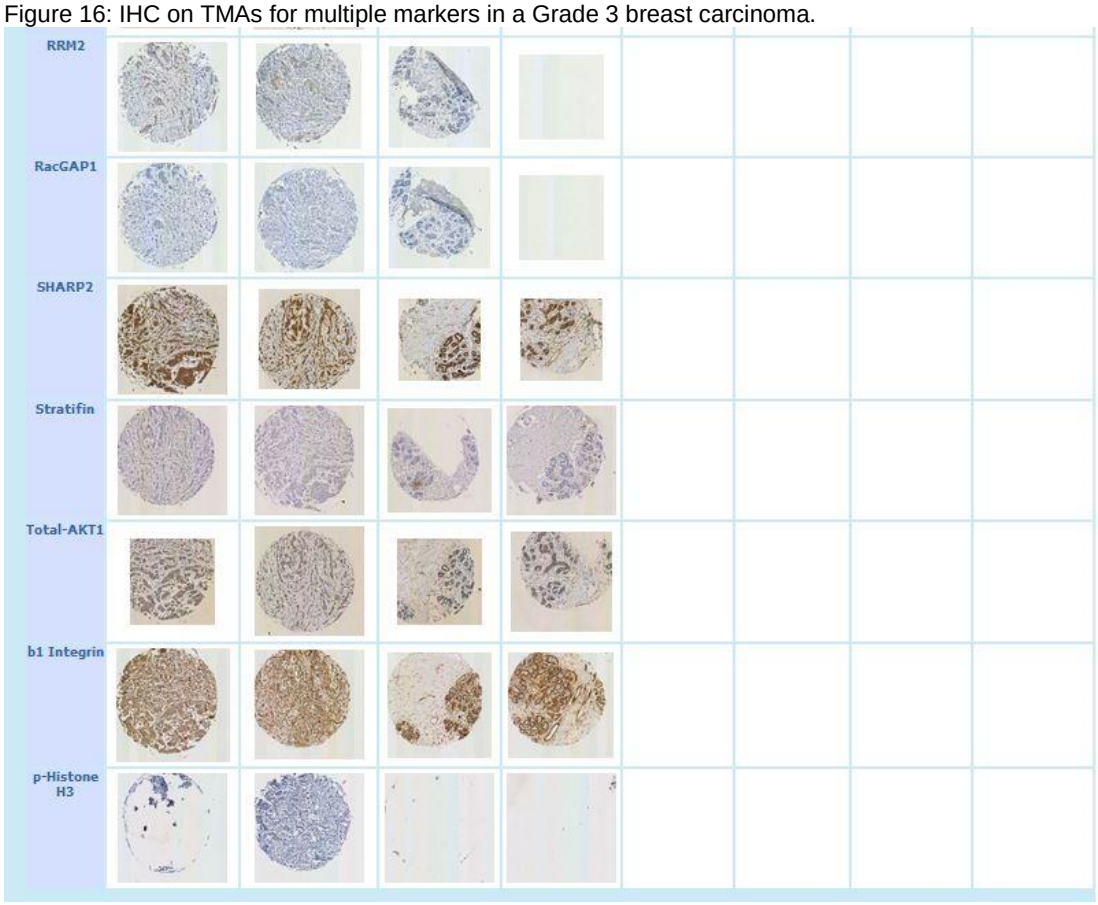


Figure 16: IHC on TMAs for multiple markers in a Grade 3 breast carcinoma.



13.5. Development and analysis of ANN models for prediction of lymph node status.

13.5.1. Introduction

NeuralWorks Predict® and NeuralSight® were used to develop a classification algorithm for the presence of lymph node metastases in breast cancer based on characteristics of the primary tumour. Variables were used that could be identified in core biopsy samples, as the primary method of breast cancer diagnosis is now a biopsy specimen. In the 1990's the most common method of initial breast carcinoma diagnosis was by frozen section, but this method would rarely be used today. Sixty-nine patients with complete datasets were used to develop the models. An additional 27 patients were used for validation.
The input variables used for the development of the models are listed in Table 18.

Table 10: Input variables for development of ANN models.

Input Variable	Criteria/Localisation of IHC
Age	Continuous
Tumour type modified	Ductal/Lobular/Mucinous/Papillary
Grade	1/2/3
Tubule formation	1/2/3
Nuclear pleomorphism	1/2/3
Mitotic Rate Rmm2	Continuous
Tumour border	Infiltrating/Pushing
Lymphoplasmacytic infiltrate	Mild/Moderate/Marked
nm-23 H-1-C Percentage	Cytoplasmic
nm-23 H-1-C Strength	
nm23H1-Cscore	
nm-23 H-1-C percentage negative	Negative cytoplasmic
nm-23 H-1-C negative cells	
nm-23 H-1-N % Pos	Nuclear
nm-23 H-1-N Strength	
nm23H1-Nscore	
ER percentage	Nuclear
ER strength	
ERscore	
PR percentage	Nuclear
PR strength	
PRscore	
c-erbB-2m percentage	Membrane
c-erbB-2m strength	
c-erbB-2Mscore	
c-erbB-2c percentage	Cytoplasmic

c-erbB-2c strength	
c-erbB-2Cscore	
A0485m%	Membrane
A0485m Strength	
A0485m Score	
A0485c%	Cytoplasmic
A0485c Strength	
A0485c Score	
TAB 250m%	Membrane
TAB250mStrength	
TAB 250 Score	
HER-2 status	Negative/Equivocal/Positive
p53 percentage	Nuclear
p53 strength	
p53score	
bcl2-N%	Nuclear
bcl2-N strength	
bcl2Nscore	
bcl2-C %	Cytoplasmic
bcl2-C strength	
bcl2Cscore	
CD9 percentage	Membrane
CD9 strength	
CD9score	
CD9 Neg%	
KAI1C %	Cytoplasmic
KAI1C strength	
KAI1Cscore	
KAI1 N % Pos	Nuclear
KAI1 N Strength	
KAI1Nscore	
KAI1 N % Neg	
Rb 1 percentage	Nuclear
Rb1 Strength	
Rb1 Score	
VEGF-C %	Cytoplasmic
VEGF-C Strength	
VEGF-C Score	
VEGF-N%	Nuclear
VEGF-N Strength	
VEGF-N Score	
p27 %	Nuclear
p27 Strength	
p27 Score	

MIB1 % F	Nuclear
MVD mm2	Continuous
p-glycoprotein C%	Cytoplasmic
p-glycoprotein C Strength	
p-glycoprotein C Score	
p-glycoprotein M%	Membrane
p-glycoprotein M Strength	
p-glycoprotein M Score	
TMA AB CA IX score	Cytoplasmic
TMA AB IHC NGFR score	Cytoplasmic
TMA AB Osteopontin score	Cytoplasmic
TMA AB p63 score	Nuclear
TMA AB P cadherin	Membrane
TMA AB CK14 score	Cytoplasmic
TMA AB IHC SMA	Cytoplasmic
TMA AB CK 5 & 6	Cytoplasmic
TMA AB IHC PR	Nuclear
TMA AB IHC ER	Nuclear
TMA AB IL8 Score	Cytoplasmic
TMA AB IL8 Stroma	Cytoplasmic
TMA AB IL8 Endothelium	Cytoplasmic
TMA AB 1SSB Score	Nuclear
TMA AB 1SSB Cytoplasm	Cytoplasmic
TMA AB chk2	Nuclear
TMA AB chk2 Cytoplasm	Cytoplasmic
TMAAB HER2 (4B5)	Membrane
TMA AB Cyclin E Score	Nuclear
TMA AB Cyclin E Membrane	Membrane
TMA AB Cyclin D1 Score	Nuclear
TMA AB Cyclin D1 Cytoplasmic	Cytoplasmic
TMA AB c-met	Membrane
TMA AB S100A4 score	Cytoplasmic
TMA AB CEP55 Intensity	Cytoplasmic
TMA AB AB Crystallin Intensity	Cytoplasmic
TMA AB AB Crystallin Percentage	
TMA AB AB Crystallin Score	
TMA SCL7A5	Cytoplasmic
TMA p53	Nuclear
TMA NDRG1	Cytoplasmic
TMA HTF9C	Cytoplasmic
TMA CEACAM5	Cytoplasmic
TMA NCSTN	Cytoplasmic
TMA TRIM29	Cytoplasmic
Mammostrat score	

Mammostrat category	Good/Moderate/Poor

13.5.2. Results

Morphological diagnosis was performed on whole tissue sections with some IHC performed on these sections with subsequent analysis of TMA sections for multiple markers. Examples of results for IHC staining on whole tissue sections are illustrated in Figures 17 – 34.

Figure 17: IHC for oestrogen receptor showing weak nuclear staining in an infiltrating lobular carcinoma (Magnification 8X).

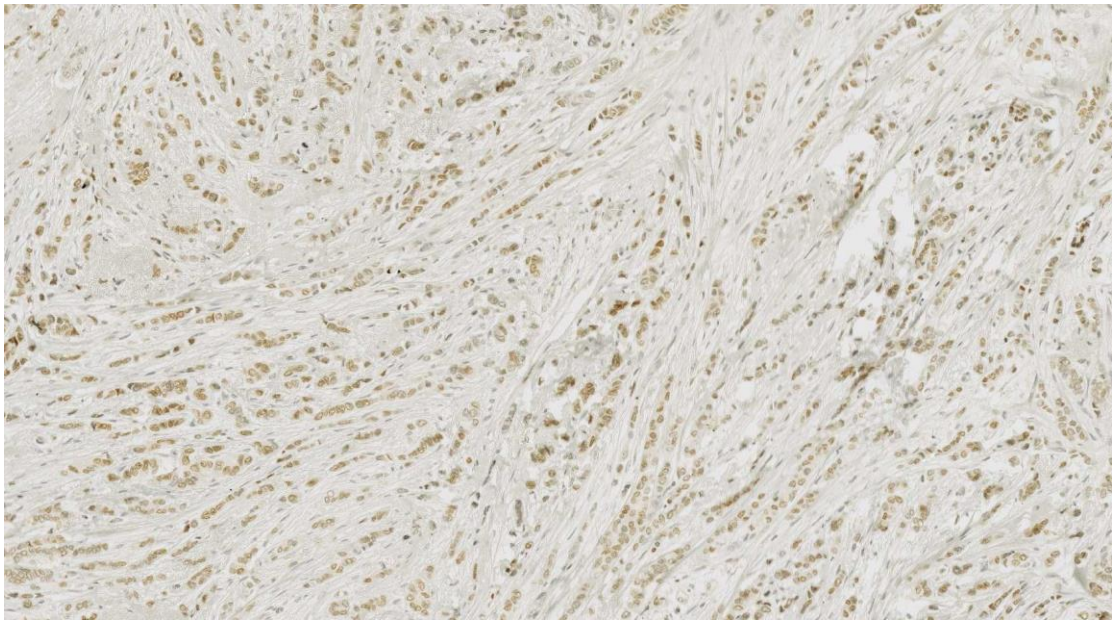


Figure 18: IHC for oestrogen receptor showing strong nuclear staining in an infiltrating ductal carcinoma (Magnification 10X).

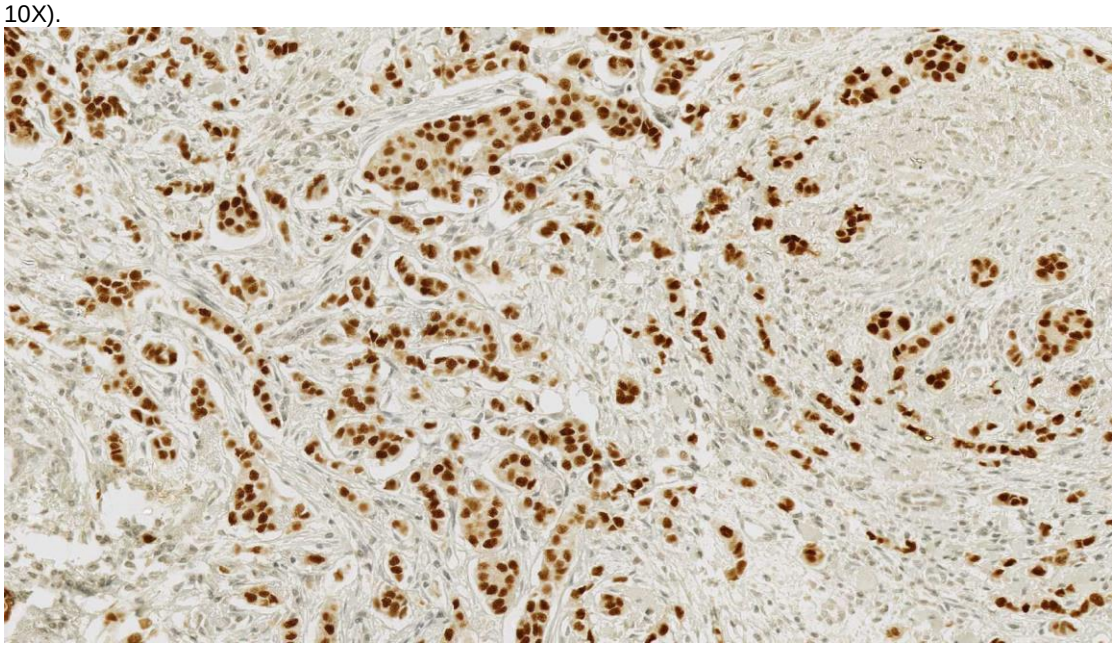


Figure 19: IHC for oestrogen receptor showing focal nuclear staining in an infiltrating ductal carcinoma (Magnification 10X).

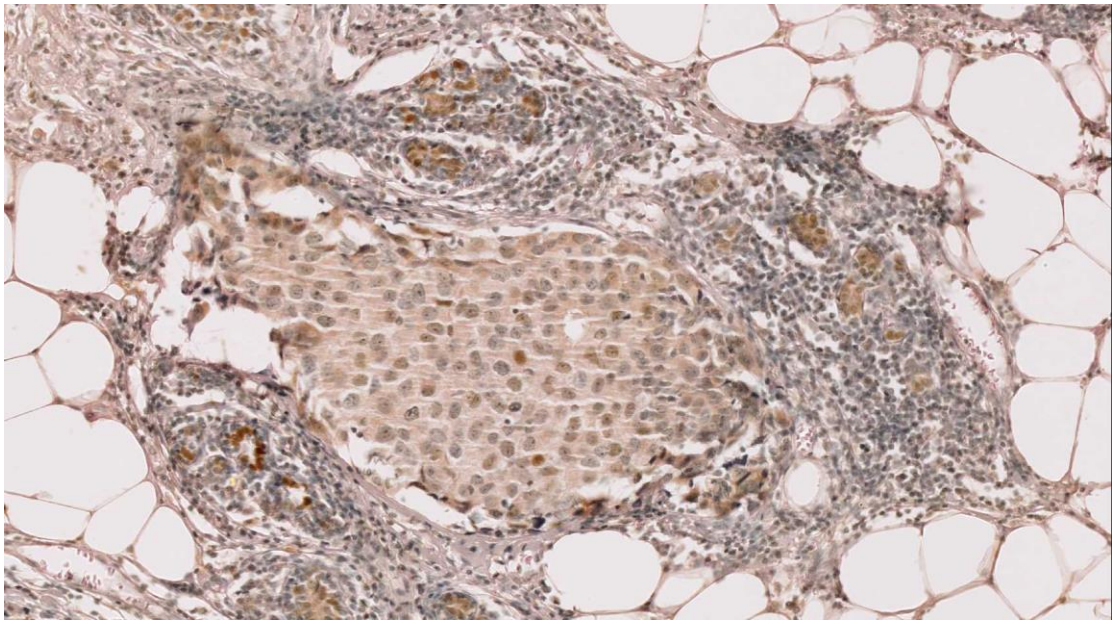


Figure 20: IHC for oestrogen receptor showing no nuclear staining in an infiltrating ductal carcinoma (Magnification 8X).

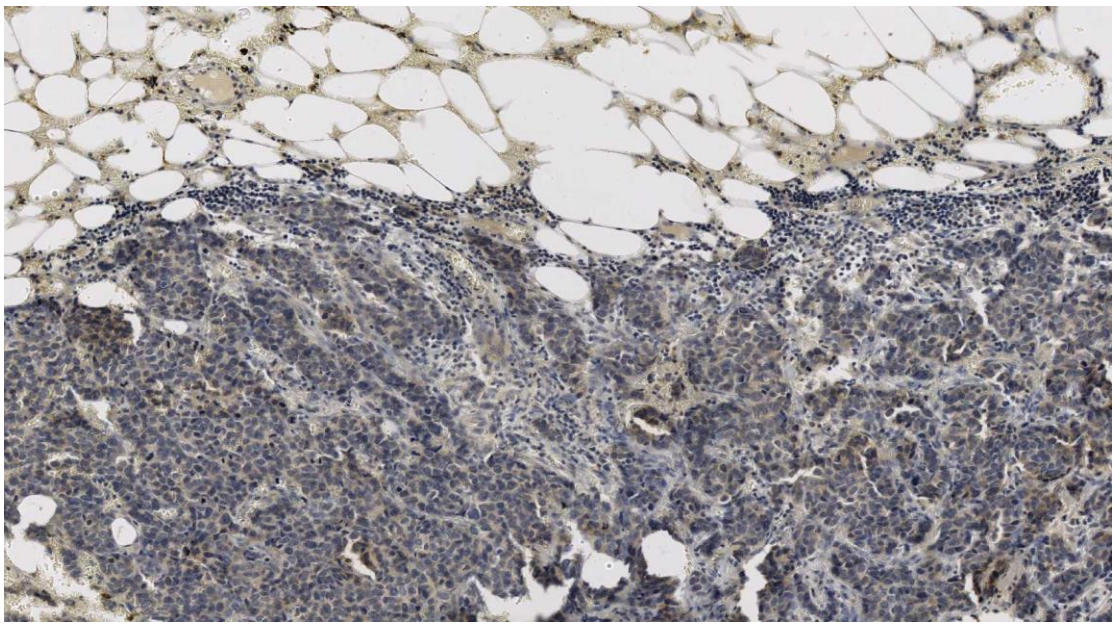


Figure 21: IHC for nm23H1 showing weak cytoplasmic staining in an infiltrating ductal carcinoma (Magnification 8X).

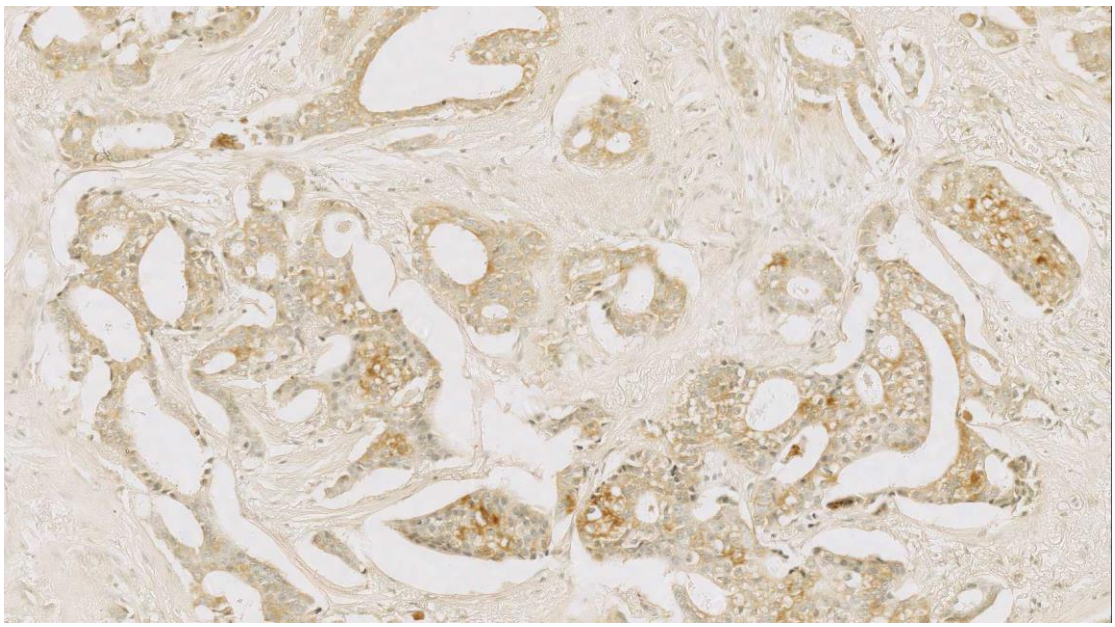


Figure 22: IHC for nm23H1 showing strong cytoplasmic staining in an infiltrating ductal carcinoma (Magnification 8X).

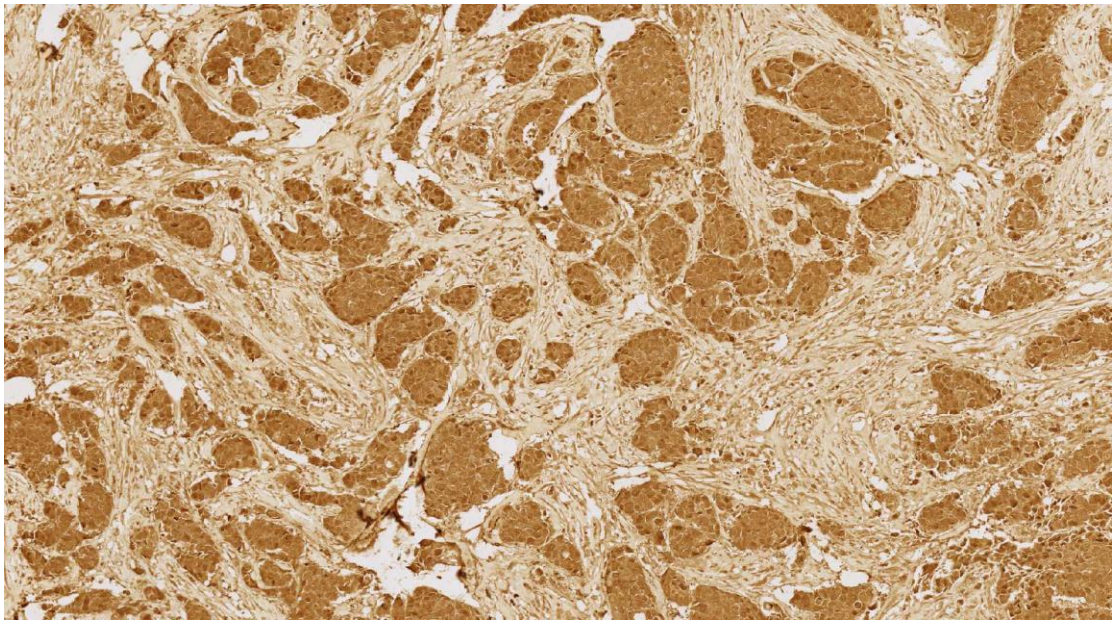


Figure 23: IHC for VEGF-C showing weak cytoplasmic staining in an infiltrating ductal carcinoma (Magnification 8X).

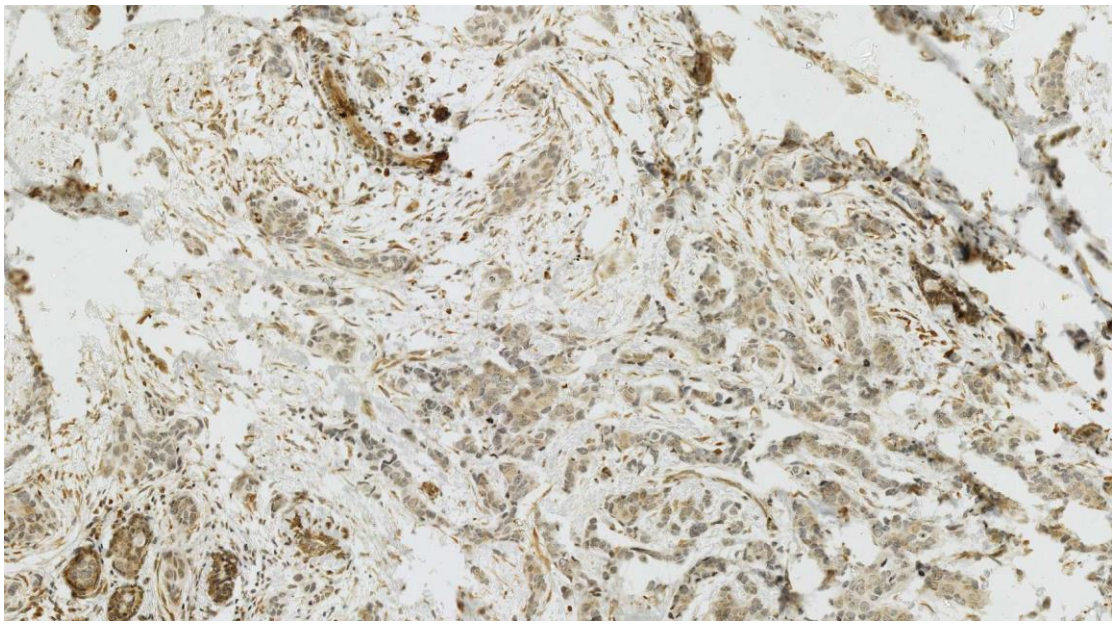


Figure 24: IHC for VEGFC showing strong cytoplasmic staining in an infiltrating lobular carcinoma (Magnification 8X).

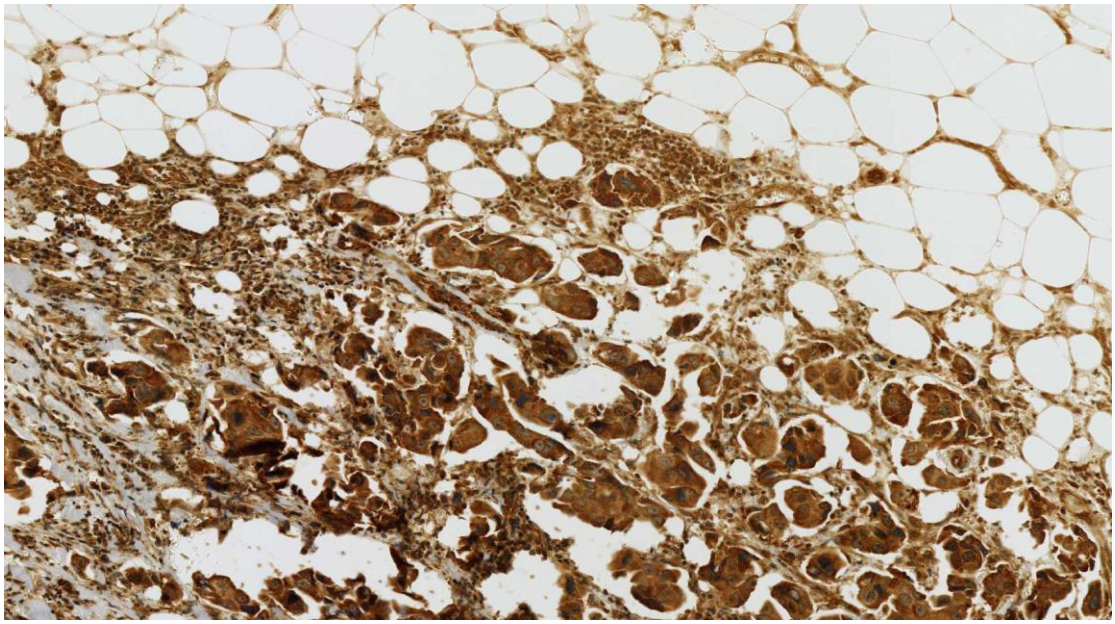


Figure 25: IHC for HER2 (Clone TAB250) showing strong membrane staining in an infiltrating ductal carcinoma (Magnification 10X). A normal breast lobule with negative staining is illustrated on the right.

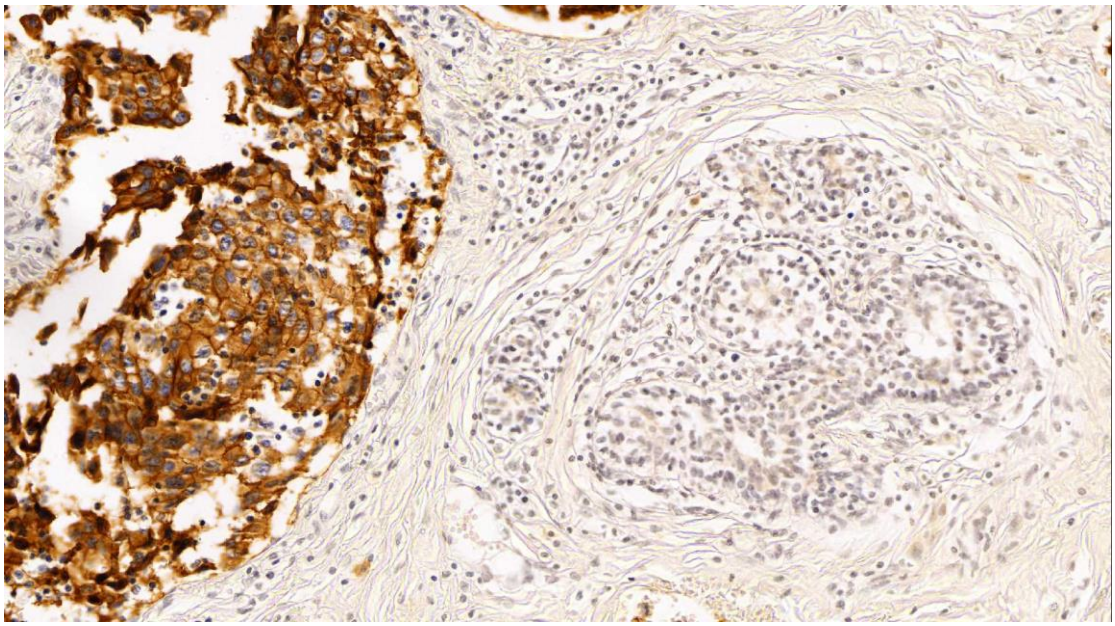


Figure 26: IHC for HER2 (Polyclonal A0485) showing strong membrane staining in the same infiltrating ductal carcinoma (Magnification 10X) as in Figure 53.

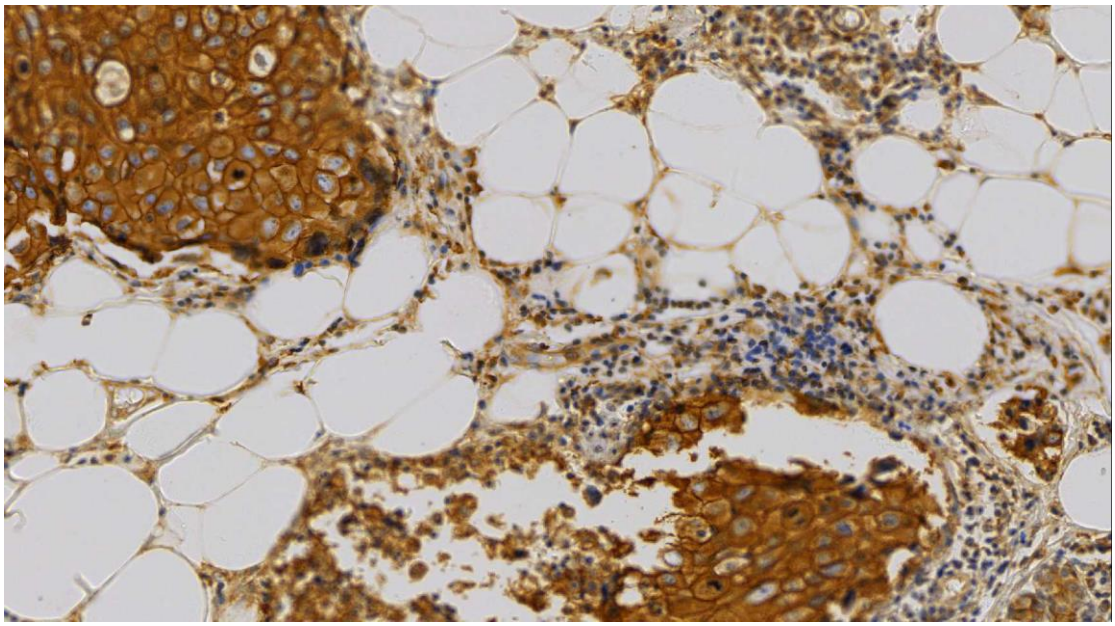


Figure 27: IHC for CD9 showing strong membrane staining in an infiltrating ductal carcinoma (Magnification 8X).

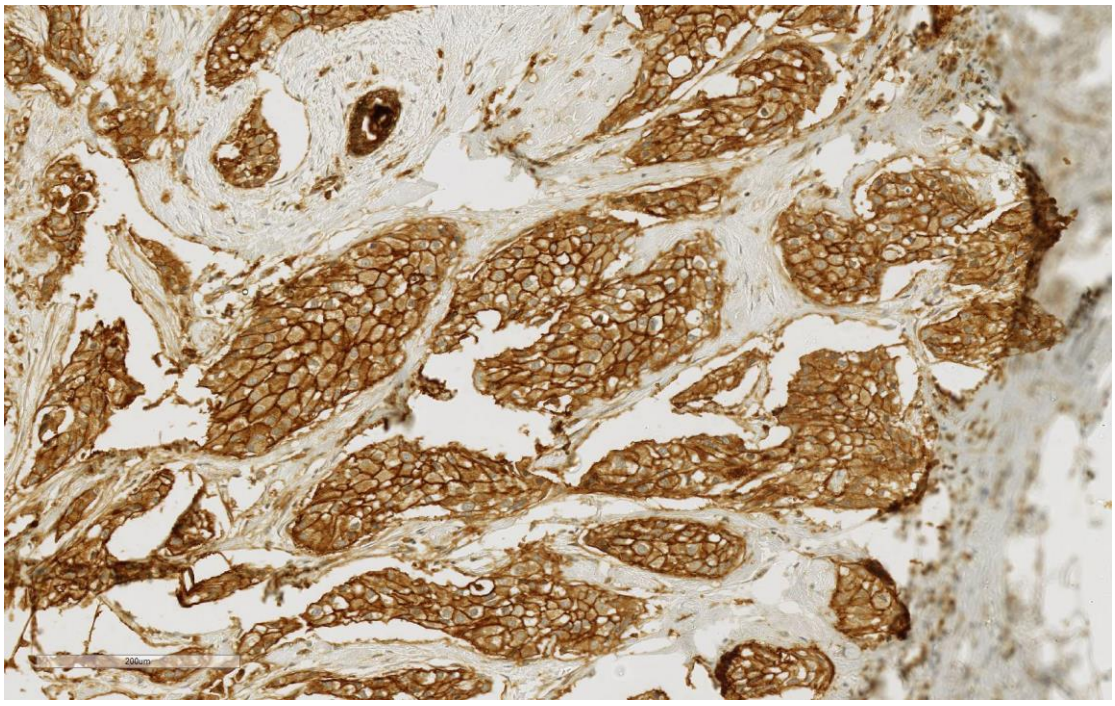


Figure 28: IHC for p27 showing strong staining in an infiltrating ductal carcinoma (Magnification 8X).

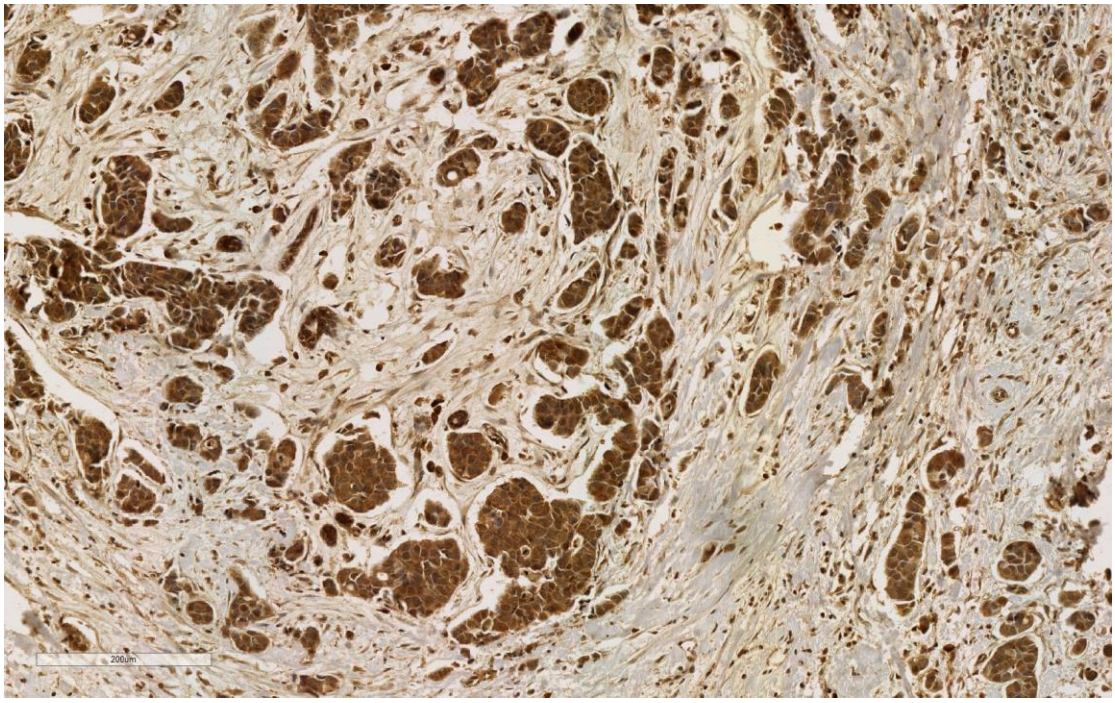


Figure 29: IHC for p27 showing weak staining in an infiltrating ductal carcinoma (Magnification 10X).

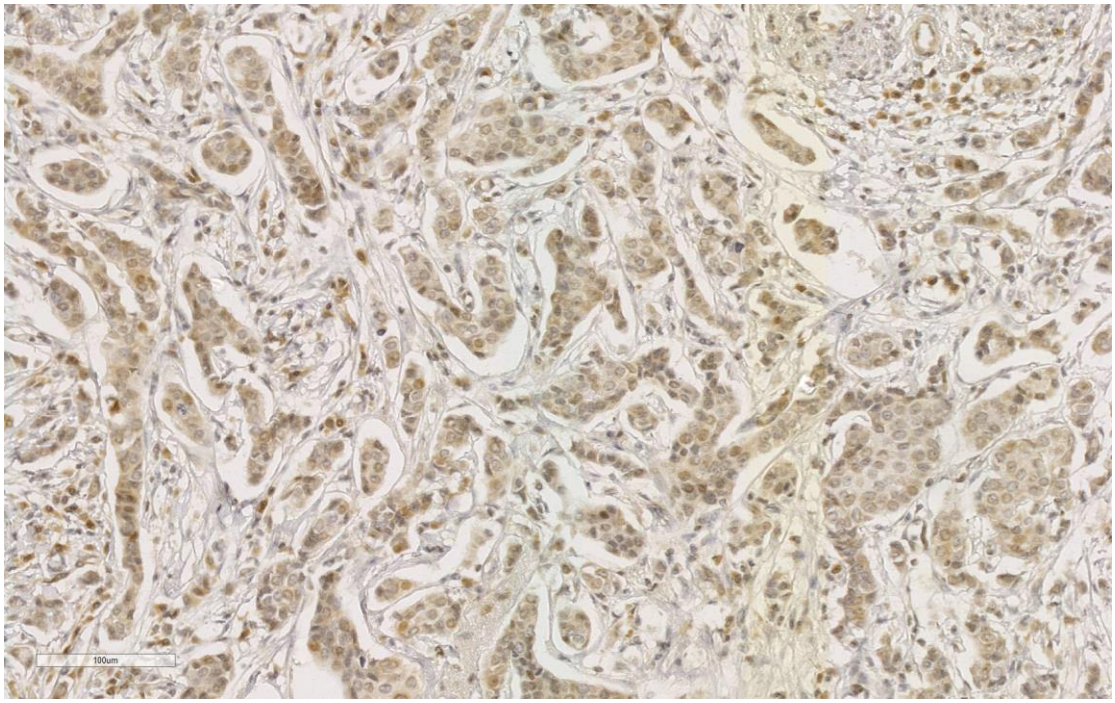


Figure 30: IHC for p27 showing negative staining in an infiltrating ductal carcinoma (Magnification 8X).

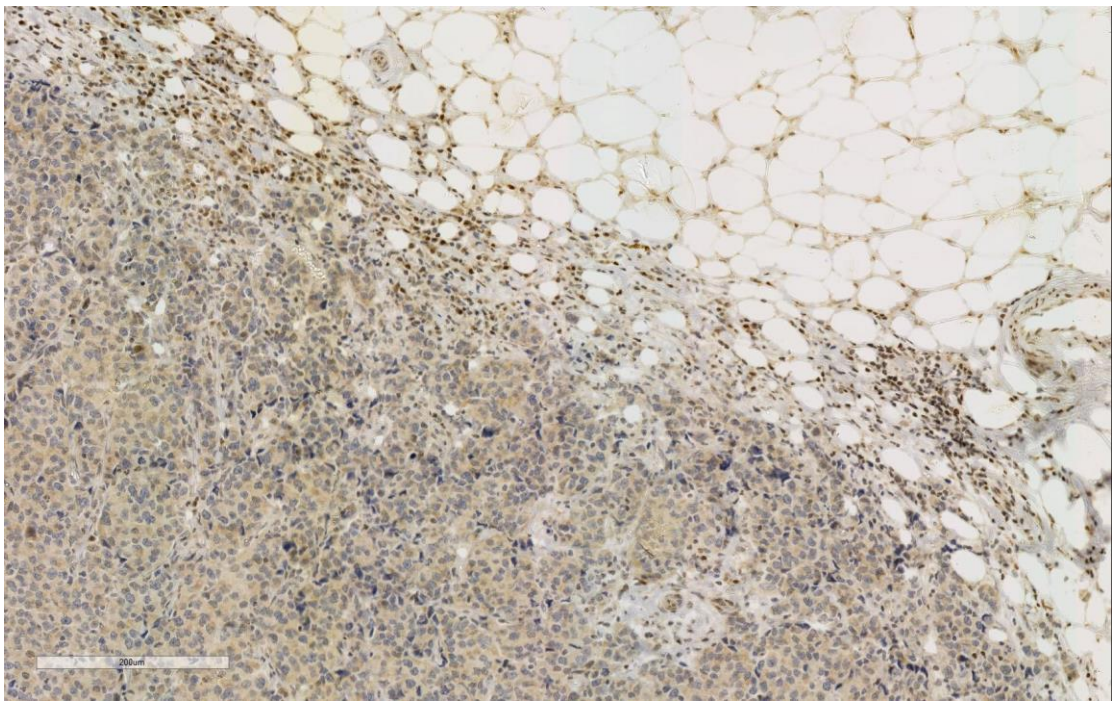


Figure 31: IHC for RB1 showing strong staining in an infiltrating ductal carcinoma (Magnification 10X).

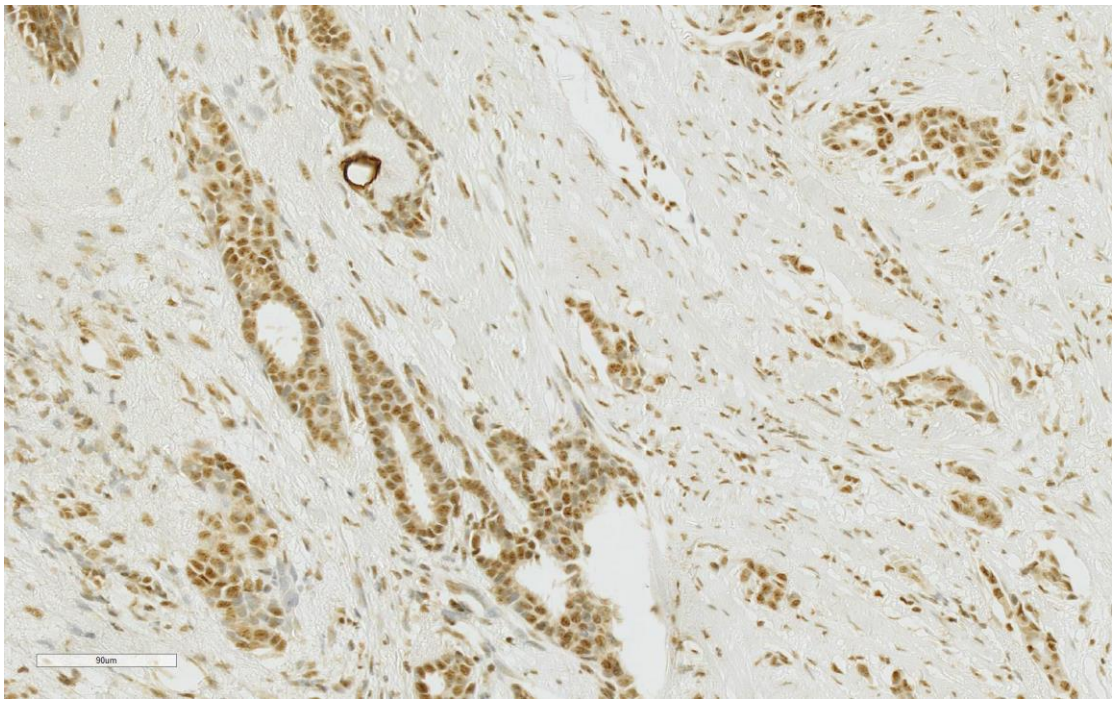


Figure 32: IHC for RB1 showing weak nuclear staining in an infiltrating ductal carcinoma (Magnification 10X).

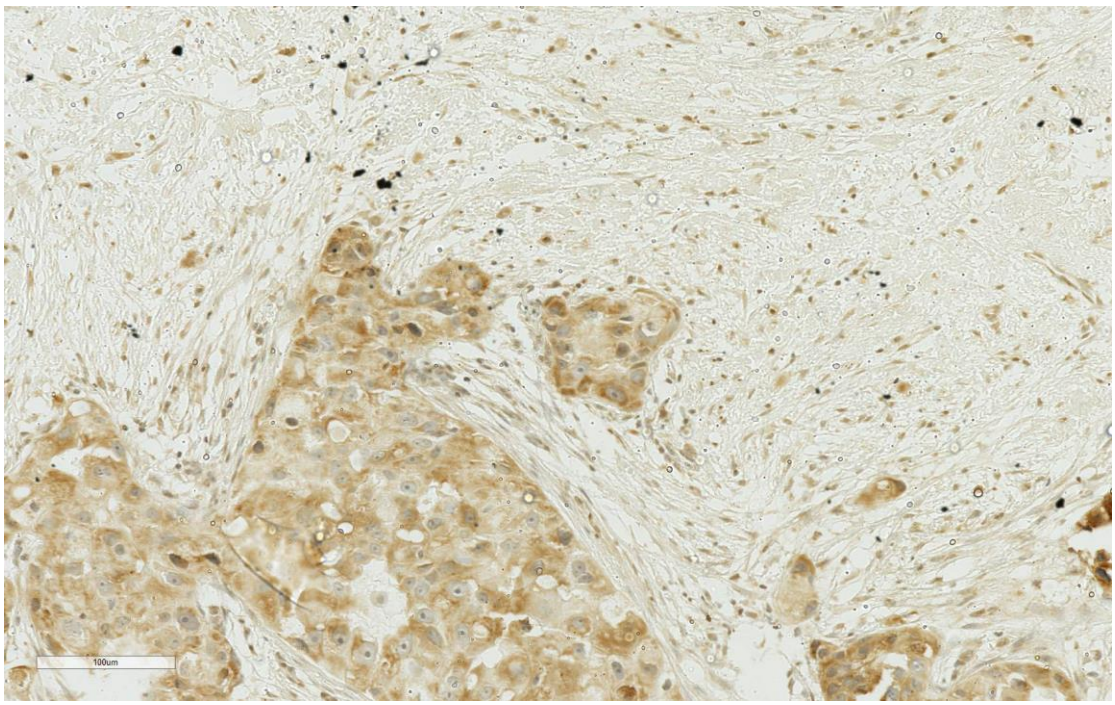


Figure 33: IHC for p-glycoprotein showing moderate cytoplasmic staining in an infiltrating ductal carcinoma (Magnification 10X).

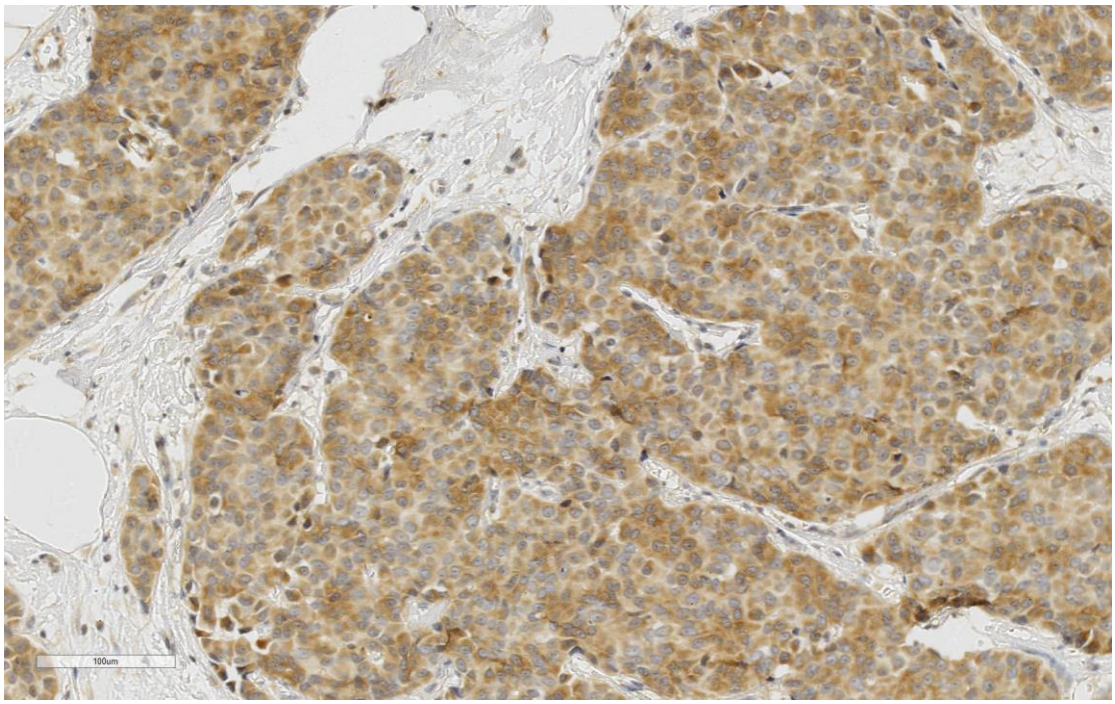
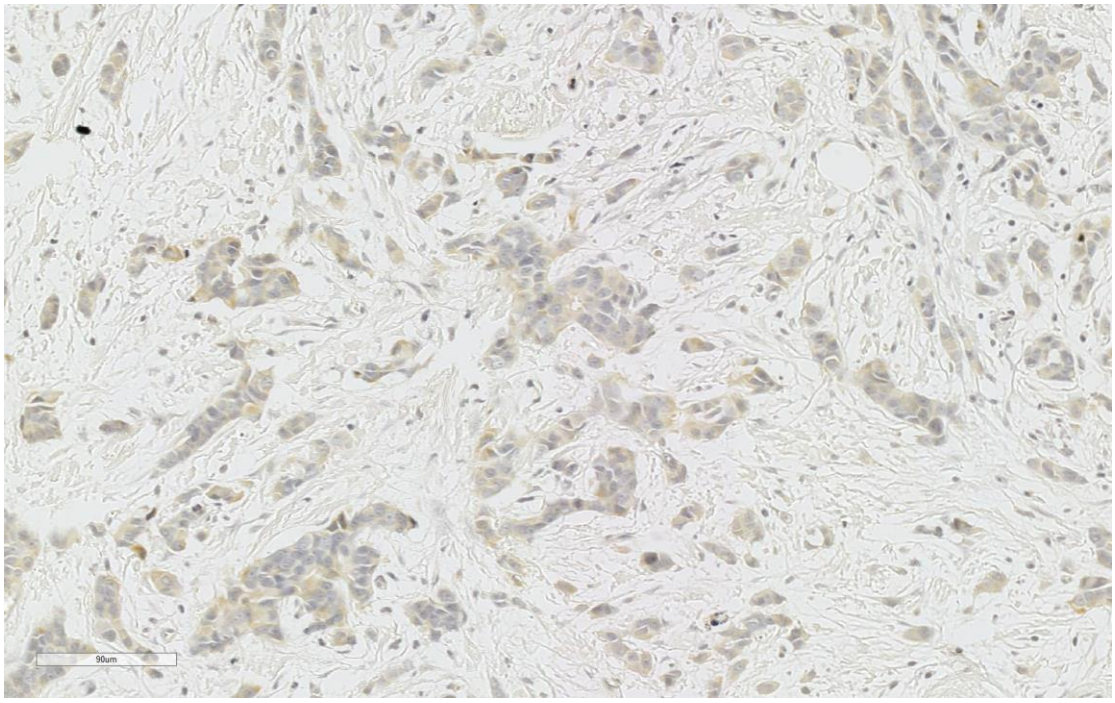


Figure 34: IHC for p-glycoprotein showing negative staining in an infiltrating ductal carcinoma (Magnification 10X).



Results of a case are illustrated in Figures 35 – 37 for multiple biomarkers using TMAs.

Figure 35: IHC on TMAs for multiple biomarkers in an illustrative breast carcinoma.

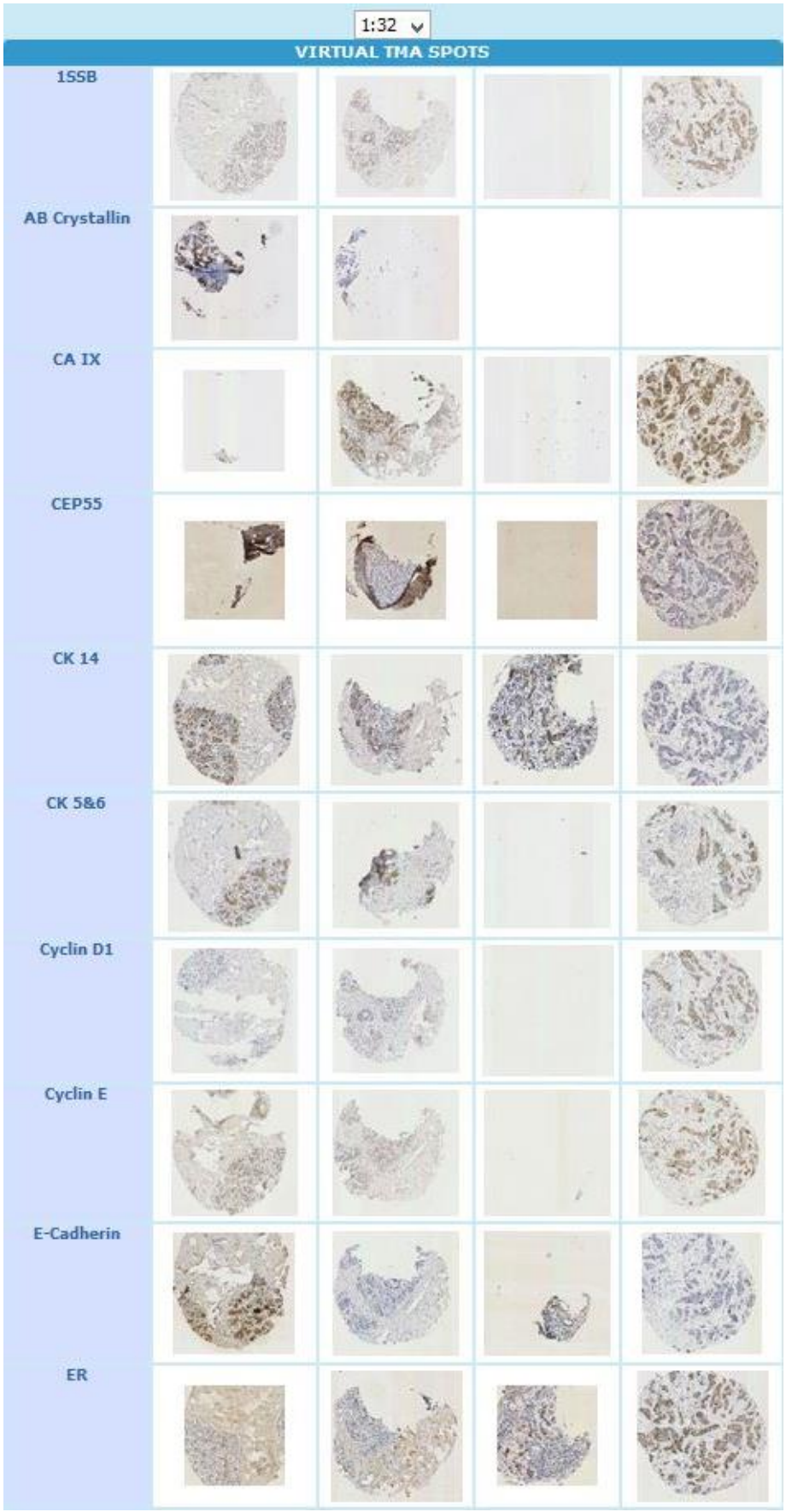


Figure 36: IHC on TMAs for multiple biomarkers in an illustrative breast carcinoma.

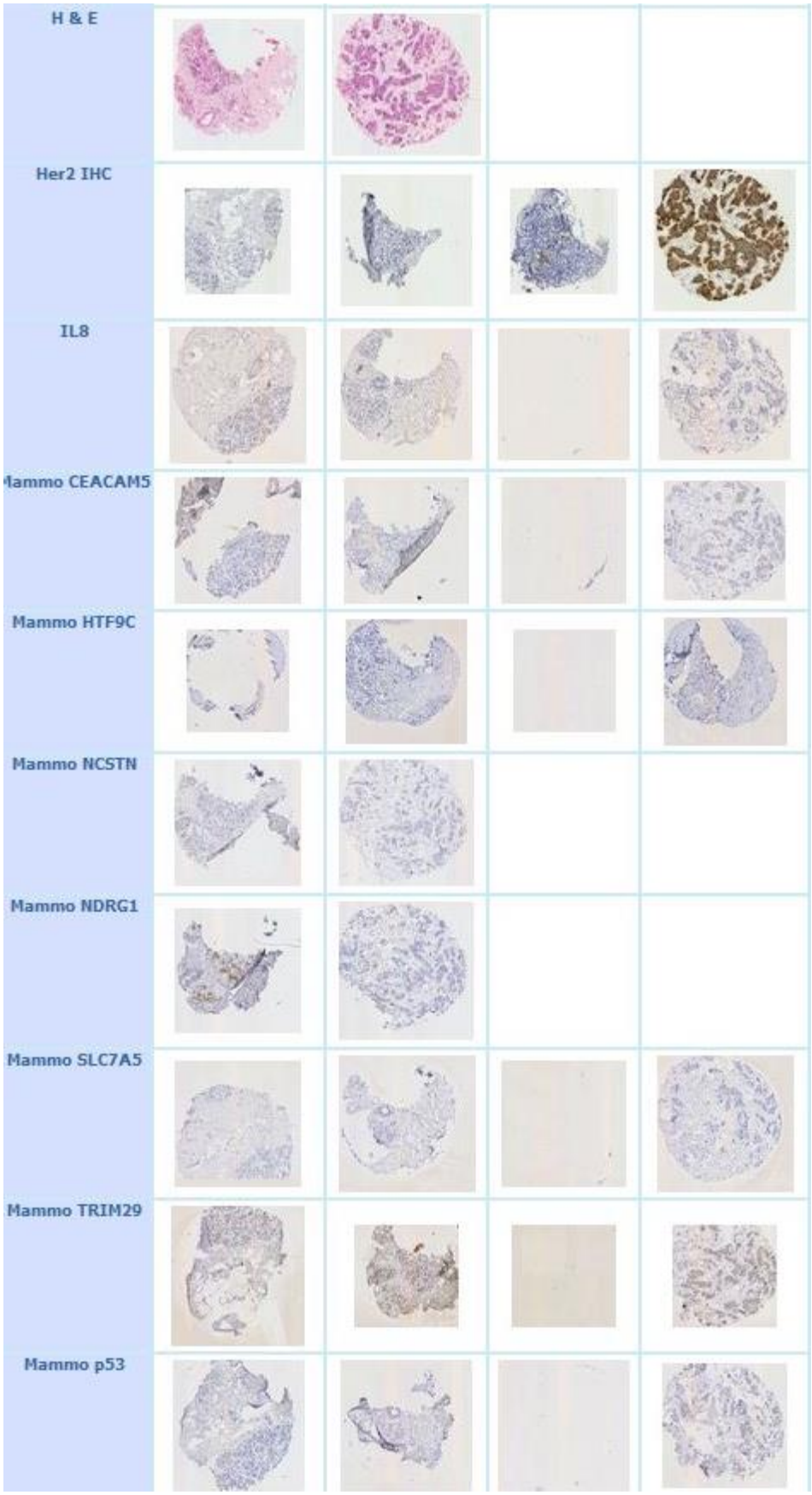
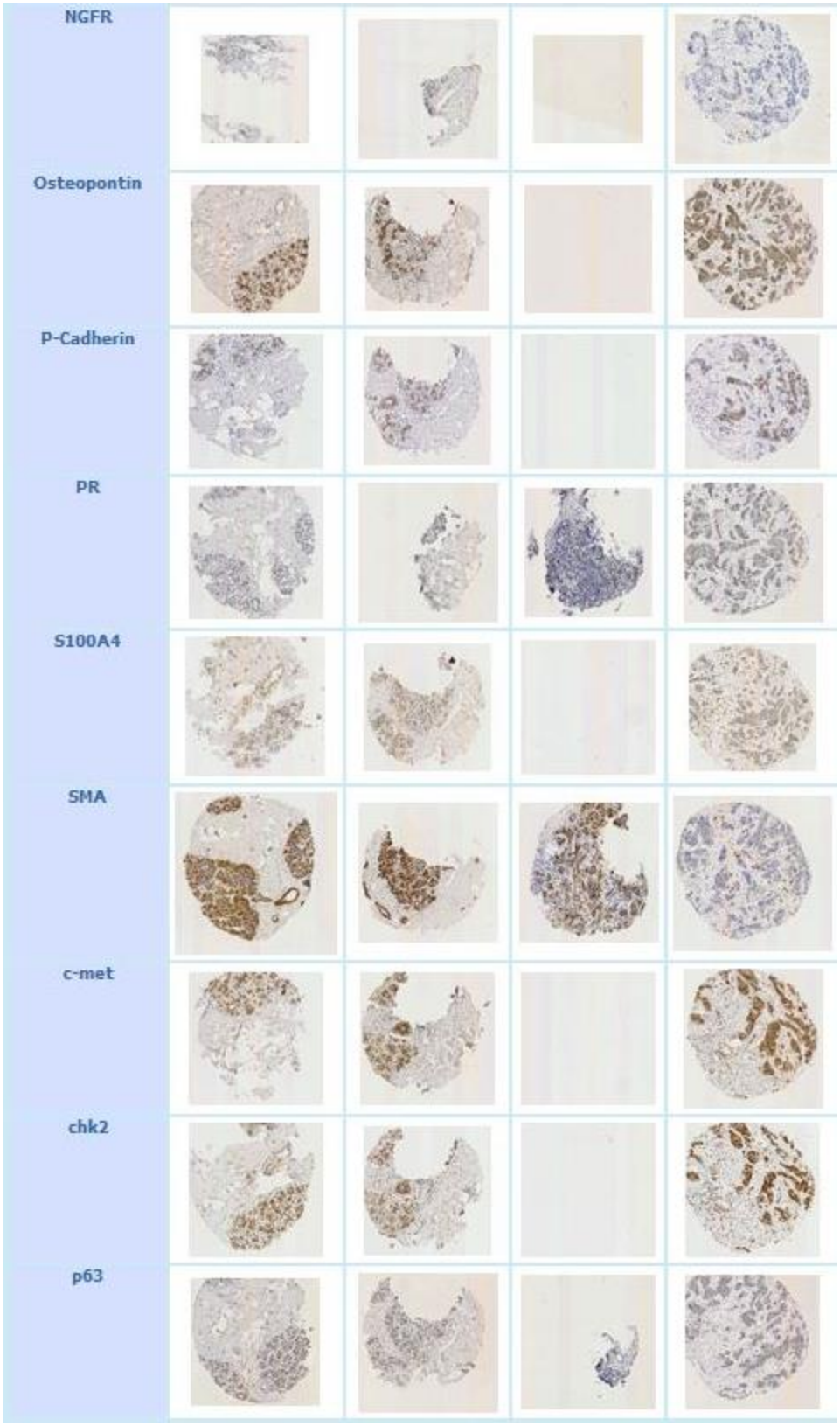


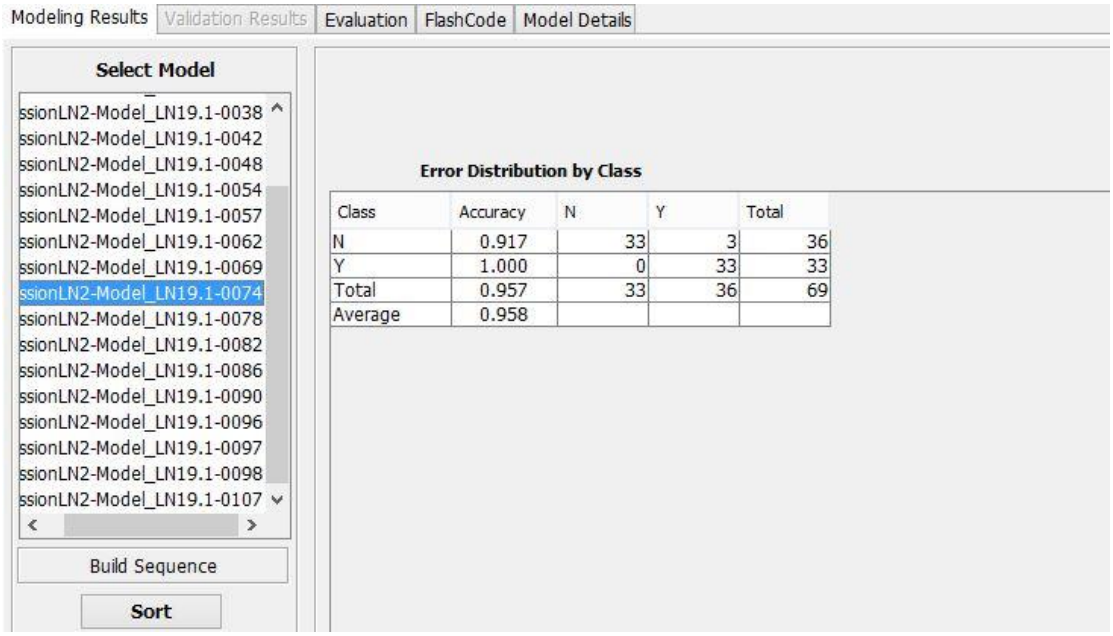
Figure 37: IHC on TMAs for multiple biomarkers in an illustrative breast carcinoma.



Multiple different models were constructed using variations in the input parameters with modifications to the ANN to include noisy data, moderately noisy data and comprehensive and exhaustive variable selection.

Model 74 showed a Sensitivity of 91.7%, a Specificity of 100%, a Positive Predictive Value of 100% and a Negative Predictive Value of 91.7%. The overall accuracy of the model was 95.8% (Figure 38).

Figure 38: Error distribution by class for ANN Model 74.



For Model 74 a Sensitivity analysis (Figure 40) and ROC (Figure 41) are generated by NeuralSight®.

Figure 39: Sensitivity analysis for ANN Model 74.

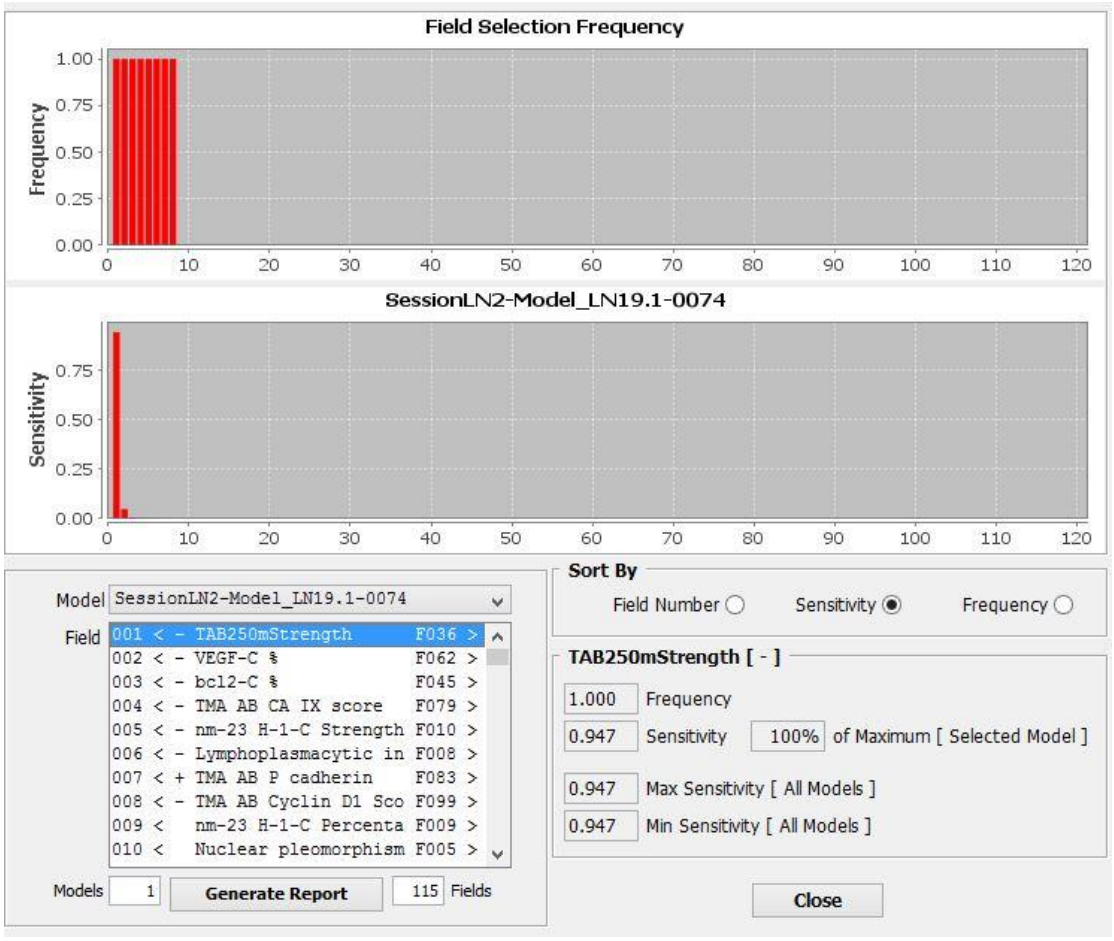
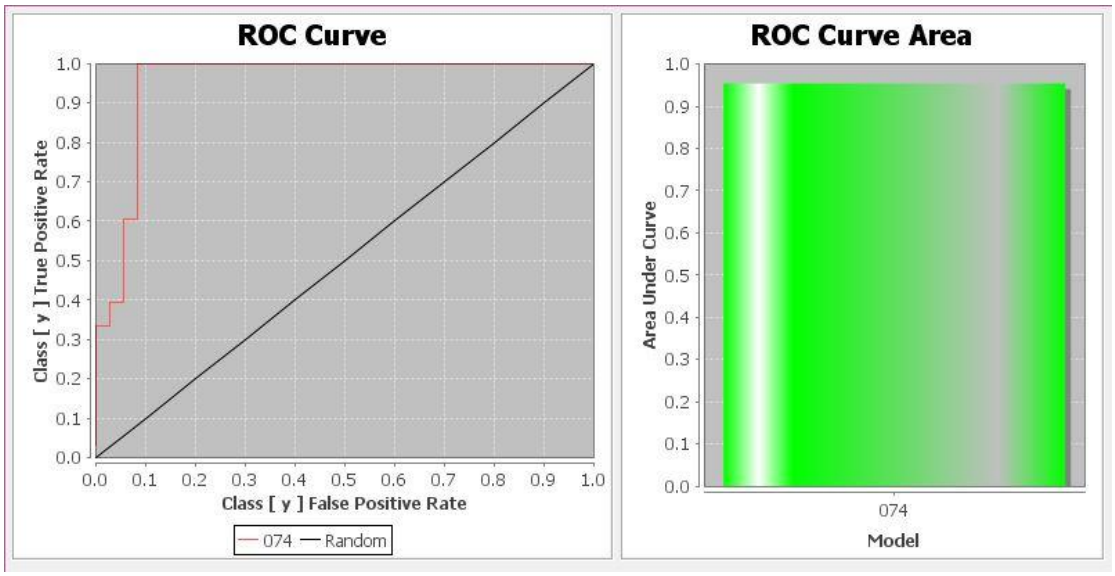
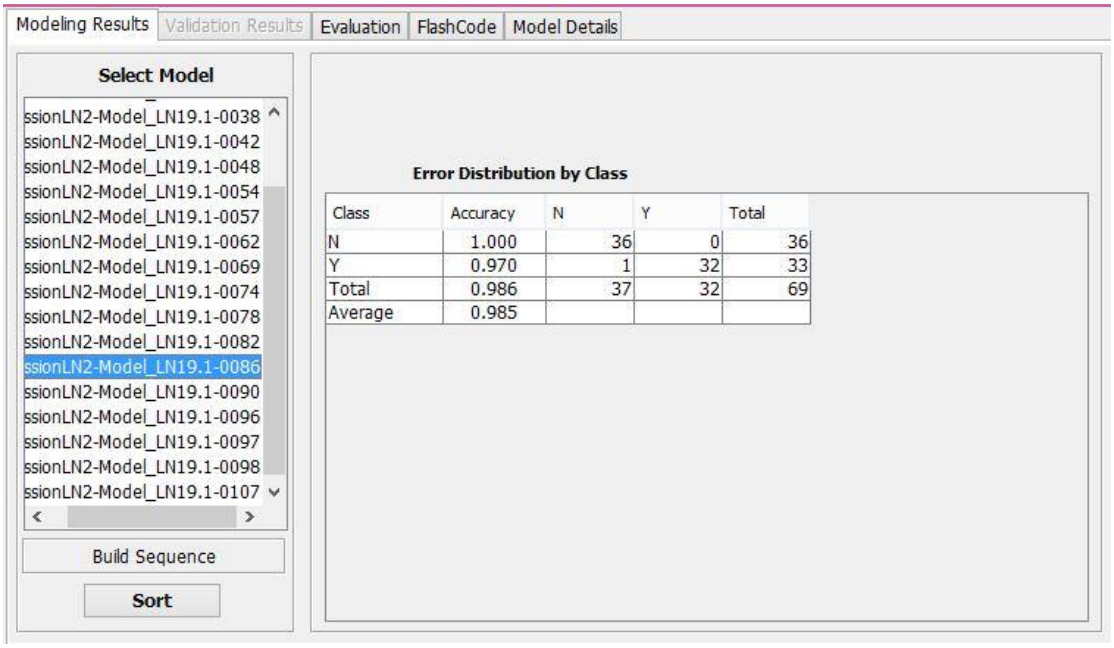


Figure 40: ROC Curve for ANN Model 74.



Model 86 showed a Sensitivity of 100%, a Specificity of 97%, a Positive Predictive Value of 97% and a Negative Predictive Value of 100%. The overall accuracy of the model was 98.5% (Figure 41).

Figure 41: Error distribution by class for ANN Model 86.



For Model 86 the Sensitivity analysis is shown in Figure 42 and ROC Curve in Figure 43.

Figure 42: Sensitivity analysis for ANN Model 86.

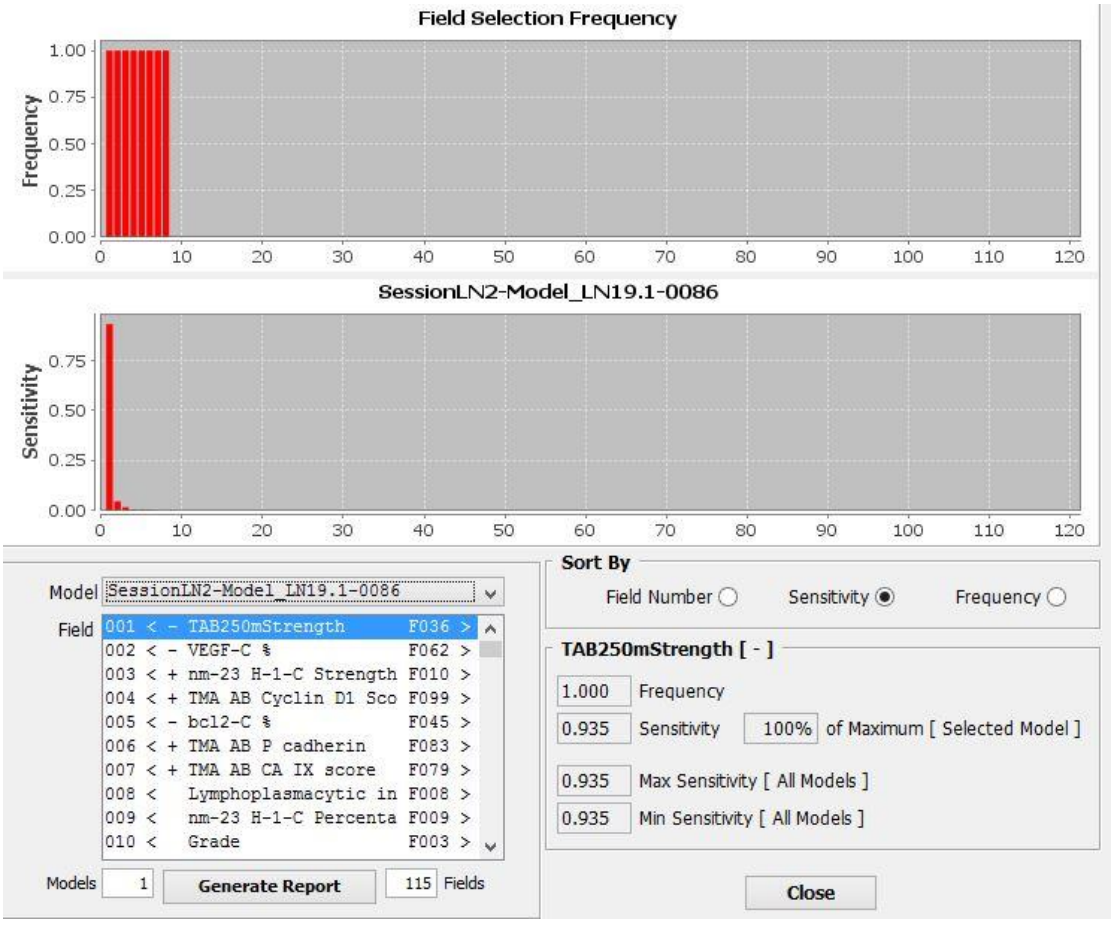
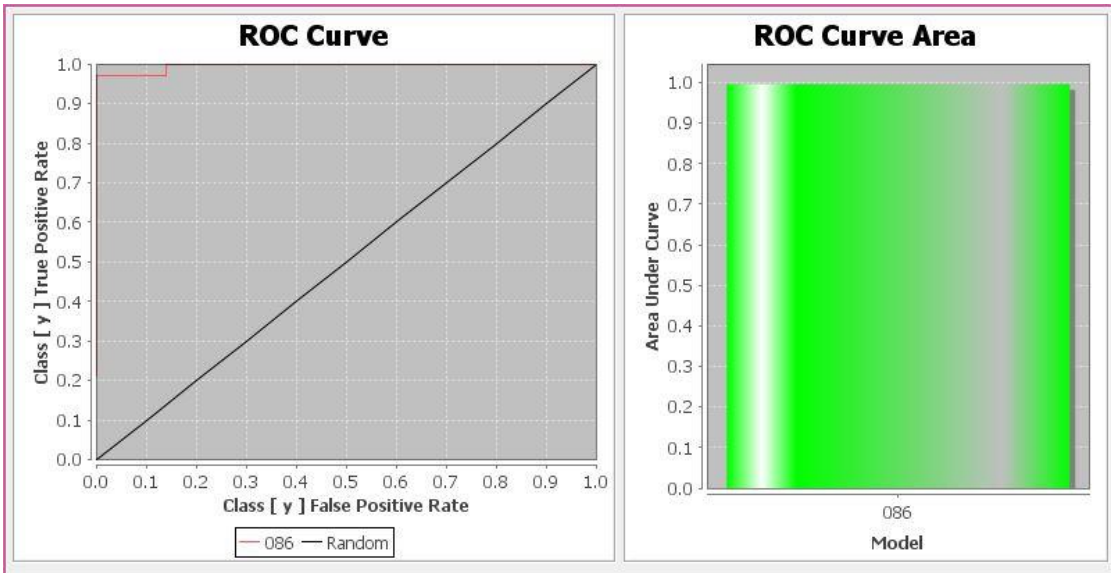


Figure 43: ROC Curve for ANN Model 86.

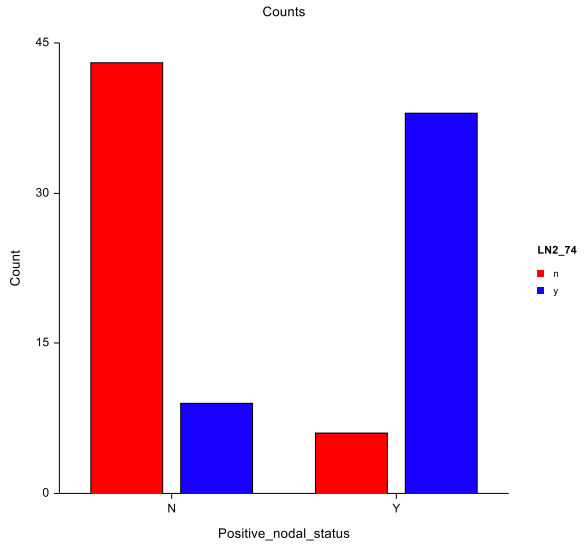


Cross tabulation was performed using the 96 patient dataset (training and validation sets). Using this combined dataset with ANN Model 74 (Table 11, Figure 44), the Sensitivity is 80.9%, the Specificity is 87.8%, the Positive Predictive Value is 86.4% and the Negative Predictive Value is 82.7%.

Table 11: Cross Tabulation table for ANN model 74 for 96 patients.

Counts Table			
	LN2_74		
Positive_nodal_status			
	n	y	Total
N	43	9	52
Y	6	38	44
Total	49	47	96

Figure 44: Plot of positive lymph node status for Model LN_74.

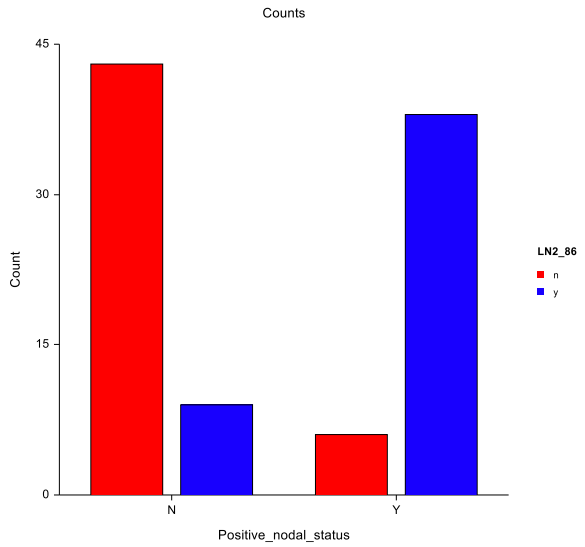


ANN Model 86 showed the same results (Table 12, Figure 45), with a Sensitivity of 80.9%, a Specificity of 87.8%, a Positive Predictive Value of 86.4% and a Negative Predictive Value of 82.7%.

Table 12: Cross Tabulation table for ANN model 86 for 96 patients.

Counts Table			
	LN2_86		
Positive_nodal_status			
	n	y	Total
N	43	9	52
Y	6	38	44
Total	49	47	96

Figure 45: Plot of positive lymph node status for Model LN_86.

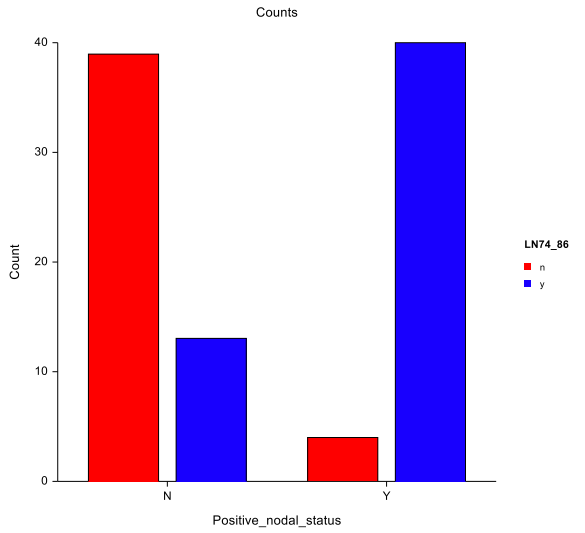


Data for both models was combined with a predicted positive nodal status being defined as called positive with either ANN Model 74 and/or ANN Model 86 (Table 13, Figure 46).

Table 13: Plot of positive lymph node status for combined Model LN74_86.

Counts Table			
	LN74_86		
Positive_nodal_status			
	n	y	Total
N	39	13	52
Y	4	40	44
Total	43	53	96

Figure 46: Plot of positive lymph node status for Model LN74_86.



The combined data from both the ANN Model 74 and ANN Model 86 showed a Sensitivity of 75.5%, a Specificity of 90.7%, a Positive Predictive Value of 90.9% and a Negative Predictive Value of 75%.

13.5.3. Discussion

Whilst the ANN was able to generate models that accurately predicted lymph node status from characteristics of the primary tumour, the models failed to generalise for use in a separate validation data set. Different parameters were tried with similar results. NeuralWorks Predict® was also used with different input variables. NeuralWorks Predict® is able to accommodate incomplete dataset entries for individual patients and models were evaluated to cover additional variables (Table 14).

Table 14: Additional variables used for development of ANN models.

Input Variable	Localisation
Tumour vol macro	
Max Macro	
Macro Vol	
Tumour vol micro	
Max Micro	
Micro Vol	
Tumour border	
Satellite lesion	
Multifocal	
Multicentric	
Lymphoplasmacytic infiltrate	
Lymphatic invasion	
Vascular invasion	
Nipple involvement-lymphatic	
Nipple - DCIS	
Nipple-Stromal infiltration	
Pagets	
p120 %	Membrane
p120 Strength	
p120 Score	
catenin %	Membrane
catenin strength	
catenin Score	
E-cad percent	Membrane
E-cad strength	
E-cad Score	

A number of models used TAB250 membrane strength as the dominant variable for prediction of lymph node status (Figure 39 and Figure 42). TAB250 is an antibody used for

IHC that binds to the extracellular domain of the HER2 protein. This correlates with the worse prognosis associated with HER2-positive breast cancer patients.
No model was able to predict lymph node status with sufficient accuracy to enable consideration of implementation into clinical practice.

13.6. Development and analysis of ANN models for prediction of breast cancer specific survival in lymph node negative patients.

13.6.1. Introduction

NeuralWorks Predict® and NeuralWare Professional II Plus® were used to develop a prediction algorithm for breast cancer specific survival in lymph node negative breast cancer patients. Time was used as a variable in the program input variables. Table 15 shows the input variables used for the prediction model. A total of 138 patients were used in the model development with stratification of the samples into training, testing and validation sets.

Table 15: Input variables used for ANN.

Hormonal Status
Age
Tumour type
Grade
Tubule formation
Nuclear pleomorphism
Mitotic Rate Rmm2
Max Micro
Micro Vol
Tumour border
Satellite lesion
Multifocal
Multicentric
Lymphoplasmacytic infiltrate
Resections margins
Lymphatic invasion
Vascular invasion
Nipple involvement-lymphatic
Nipple-Stromal infiltration
Pagets
Skin involvement-lymphatic
Skin Involvement-stroma
Number of Positive Nodes
Total Number of Nodes
Positive nodal status
Apical node involvement
ExtraNodal spread
Micrometastases
Macrometastases
nm-23 H-1-C Percentage
nm-23 H-1-C Strength
nm23H1-Cscore
nm-23 H-1-C percentage negative
nm-23 H-1-C negative cells

nm-23 H-1-N % Pos
nm-23 H-1-N Strength
nm23H1-Nscore
ER percentage
ER strength
ERscore
PR percentage
PR strength
PRscore
c-erbB-2m percentage
c-erbB-2m strength
c-erbB-2Mscore
c-erbB-2c percentage
c-erbB-2c strength
c-erbB-2Cscore
A0485m%
A0485m Strength
A0485m Score
A0485c%
A0485c Strength
A0485c Score
p53 percentage
p53 strength
p53score
bcl2-N%
bcl2-N strength
bcl2Nscore
bcl2-C %
bcl2-C strength
bcl2Cscore
CD9 percentage
CD9 strength
CD9score
CD9 Neg%
KAI-1C percentage
KAI-1C strength
KAI-1Cscore
KAI-1 C % negative
KAI-1 N % Pos
KAI-1 N Strength
KAI-1Nscore
KAI-1 N % Neg
TAB 250m%
TAB250mStrength
TAB 250 Score
Rb 1 percentage
Rb1 Strength
Rb1 Score
VEGF-C %
VEGF-C Strength
VEGF-C Score

VEGF-N%
VEGF-N Strength
VEGF-N Score
p27 %
p27 Strength
p27 Score
MIB1 % F
MVD mm2
p-glycoprotein C%
p-glycoprotein C Strength
p-glycoprotein C Score
p-glycoprotein M%
p-glycoprotein M Strength
p-glycoprotein M Score
Palpability
Surgical Treatment
Breast reconstruction
Combination chemotherapy
Toremifene
Tamoxifen
Highdose chemo
Ovarian ablation
Radiotherapy pre mastectomy
Local radiotherapy after lumpectomy
Post mastectomy XRT

13.6.2. Results

A number of models were developed to enable prediction of BCSS for individual patients. The accuracy of the models is indicated in Table 16.

Table 16: Analysis of ANN model for prediction of BCSS.

Model 1								
Death from disease	R	Net-R	Avg. Abs.	Max. Abs.	RMS	Accuracy (20%)	Conf. Interval (95%)	Records
All	0.9683189	0.9683189	0.03992268	0.9918341	0.09329377	0.9681471	0.1813712	9952
Train	0.971567	0.971567	0.03890297	0.9788387	0.0898455	0.9696698	0.1746878	5572
Test	0.9625783	0.9625783	0.04248443	0.9918341	0.1002246	0.9636895	0.1949362	2396
Valid	0.9653994	0.9653994	0.03969276	0.9530491	0.09411453	0.969254	0.1830754	1984
Model 4								
Death from disease	R	Net-R	Avg. Abs.	Max. Abs.	RMS	Accuracy (20%)	Conf. Interval (95%)	Records
All	0.9701439	0.9701439	0.04322206	0.9858124	0.08961993	0.9756913	0.1742292	9873
Train	0.9757502	0.9757502	0.04155975	0.8926328	0.08184663	0.9782805	0.1591358	5525
Test	0.9656347	0.9656347	0.04382167	0.9858124	0.09605615	0.9735738	0.1868291	2384
Valid	0.9589324	0.9589324	0.04717053	0.985256	0.1016397	0.9709776	0.1977153	1964
Model 5								
Death from	R	Net-R	Avg. Abs.	Max. Abs.	RMS	Accuracy (20%)	Conf. Interval (95%)	Records

disease								
All	0.9701439	0.9701439	0.04322206	0.9858124	0.08961993	0.9756913	0.1742292	9873
Train	0.9757502	0.9757502	0.04155975	0.8926328	0.08184663	0.9782805	0.1591358	5525
Test	0.9656347	0.9656347	0.04382167	0.9858124	0.09605615	0.9735738	0.1868291	2384
Valid	0.9589324	0.9589324	0.04717053	0.985256	0.1016397	0.9709776	0.1977153	1964
Model 6								
Death from disease	R	Net-R	Avg. Abs.	Max. Abs.	RMS	Accuracy (20%)	Conf. Interval (95%)	Records
All	0.9562074	0.9562074	0.05449945	0.901729	0.1076865	0.9564469	0.2093522	9873
Train	0.9604197	0.9604197	0.05269169	0.901729	0.1034222	0.960181	0.2010855	5525
Test	0.9570406	0.9570406	0.05511421	0.8827108	0.1069765	0.9584732	0.2080692	2384
Valid	0.9424571	0.9424571	0.05883866	0.8483678	0.1196406	0.9434827	0.2327314	1964
Model 7								
Death from disease	R	Net-R	Avg. Abs.	Max. Abs.	RMS	Accuracy (20%)	Conf. Interval (95%)	Records
All	0.9728041	0.9728041	0.04552943	0.9395897	0.08509292	0.9708295	0.1654282	9873
Train	0.9773541	0.9773541	0.0429435	0.9395897	0.07851739	0.9737557	0.1526627	5525
Test	0.969571	0.969571	0.04804412	0.9382062	0.09027685	0.9685403	0.1755884	2384
Valid	0.9631481	0.9631481	0.04975152	0.8516374	0.09572736	0.9653768	0.1862142	1964
Model 8								
Death from disease	R	Net-R	Avg. Abs.	Max. Abs.	RMS	Accuracy (20%)	Conf. Interval (95%)	Records
All	0.9705003	0.9705003	0.04494727	0.9773265	0.08817631	0.9659678	0.1714226	9873
Train	0.9748199	0.9748199	0.04275449	0.925543	0.08227392	0.9692308	0.1599666	5525
Test	0.9683646	0.9683646	0.04664741	0.9773265	0.09158222	0.9643456	0.1781273	2384
Valid	0.96016	0.96016	0.04905216	0.9240964	0.09930804	0.9587576	0.1931795	1964
Model 9								
Death from disease	R	Net-R	Avg. Abs.	Max. Abs.	RMS	Accuracy (20%)	Conf. Interval (95%)	Records
All	0.9643049	0.9643049	0.04941627	1.018502	0.0972417	0.9655628	0.1890466	9844
Train	0.9673396	0.9673396	0.04809762	0.9551042	0.09186941	0.9670111	0.1786233	5517
Test	0.9596448	0.9596448	0.05117679	1.018502	0.1056069	0.9622561	0.2054066	2358
Valid	0.9618613	0.9618613	0.05100269	0.9876812	0.1013435	0.9654647	0.1971386	1969
Model 10								
Death from disease	R	Net-R	Avg. Abs.	Max. Abs.	RMS	Accuracy (20%)	Conf. Interval (95%)	Records
All	0.9643328	0.9643328	0.04942197	1.018502	0.09753393	0.9652735	0.189615	9762
Train	0.9626197	0.9626197	0.04969383	1.018502	0.09917076	0.9640445	0.1928198	5479
Test	0.9657402	0.9657402	0.05011302	0.8630665	0.0965404	0.9640565	0.1877733	2337
Valid	0.9673443	0.9673443	0.04782663	0.9512262	0.09401018	0.9701953	0.1828751	1946
Model 11								
Death from disease	R	Net-R	Avg. Abs.	Max. Abs.	RMS	Accuracy (20%)	Conf. Interval (95%)	Records
All	0.9486263	0.9486263	0.04332662	0.9566832	0.08806778	0.9761346	0.1712065	12361
Train	0.9528273	0.9528273	0.04230153	0.9448596	0.0837947	0.9778966	0.1629147	6922
Test	0.9396203	0.9396203	0.04512379	0.9566832	0.09565381	0.9737108	0.186024	2967

Valid	0.9482878	0.9482878	0.04403996	0.9212241	0.09021928	0.97411	0.1754725	2472
Model 12								
Death from disease	R	Net-R	Avg. Abs.	Max. Abs.	RMS	Accuracy (20%)	Conf. Interval (95%)	Records
All	0.9442828	0.9442828	0.04331891	0.8920973	0.09278711	0.97662	0.180381	12361
Train	0.9451154	0.9451154	0.04271107	0.8859505	0.09142593	0.9773187	0.1777515	6922
Test	0.9464115	0.9464115	0.04321826	0.8796806	0.091033	0.9784294	0.1770376	2967
Valid	0.9397061	0.9397061	0.04514178	0.8920973	0.09848142	0.9724919	0.1915421	2472
Model 13								
Death from disease	R	Net-R	Avg. Abs.	Max. Abs.	RMS	Accuracy (20%)	Conf. Interval (95%)	Records
All	0.9191191	0.9191191	0.04623149	0.9562494	0.1096698	0.9631098	0.2132015	12361
Train	0.9230137	0.9230137	0.04535891	0.9562494	0.1062073	0.9648945	0.2064895	6922
Test	0.9188482	0.9188482	0.04654289	0.9525762	0.1100344	0.9615774	0.2139908	2967
Valid	0.9090645	0.9090645	0.0483011	0.9554956	0.1184246	0.9599515	0.2303307	2472
Model 14								
Death from disease	R	Net-R	Avg. Abs.	Max. Abs.	RMS	Accuracy (20%)	Conf. Interval (95%)	Records
All	0.9274038	0.9274038	0.06127508	0.9776498	0.1076563	0.9652132	0.2092873	12361
Train	0.938977	0.938977	0.06208301	0.9348407	0.1090963	0.9620733	0.2121178	5537
Test	0.9366283	0.9366283	0.06122585	0.9203587	0.110388	0.968829	0.2147051	2374
Valid	0.9234235	0.9234235	0.06331819	0.9776498	0.1125795	0.9607605	0.2189623	2472
Model 15								
Death from disease	R	Net-R	Avg. Abs.	Max. Abs.	RMS	Accuracy (20%)	Conf. Interval (95%)	Records
All	0.9611858	0.9611858	0.0545365	0.9088011	0.08536106	0.9769436	0.1659446	12361
Train	0.9628542	0.9628542	0.05386386	0.9088011	0.0835251	0.9779242	0.1623837	8652
Test	0.9575254	0.9575254	0.05610556	0.8603248	0.08949753	0.9746562	0.1740343	3709
Valid	0.9611858	0.9611858	0.0545365	0.9088011	0.08536106	0.9769436	0.1659446	12361
Model 16								
Death from disease	R	Net-R	Avg. Abs.	Max. Abs.	RMS	Accuracy (20%)	Conf. Interval (95%)	Records
All	0.9627222	0.9627222	0.03542585	0.9710088	0.07736106	0.9770245	0.1503923	12361
Train	0.9635464	0.9635464	0.03509719	0.9710088	0.07599323	0.976884	0.1477407	8652
Test	0.9608842	0.9608842	0.03619252	0.8454505	0.08046146	0.9773524	0.156463	3709
Valid	0.9627222	0.9627222	0.03542585	0.9710088	0.07736106	0.9770245	0.1503923	12361
Model 17								
Death from disease	R	Net-R	Avg. Abs.	Max. Abs.	RMS	Accuracy (20%)	Conf. Interval (95%)	Records
All	0.9537181	0.9537181	0.0451647	0.9282653	0.08414366	0.9716851	0.1635779	12361
Train	0.9534523	0.9534523	0.0451846	0.8522717	0.08470036	0.9707582	0.1646685	8652
Test	0.9543585	0.9543585	0.04511828	0.9282653	0.08283049	0.9738474	0.1610697	3709
Valid	0.9537181	0.9537181	0.0451647	0.9282653	0.08414366	0.9716851	0.1635779	12361
Model 18								
Death	R	Net-R	Avg. Abs.	Max. Abs.	RMS	Accuracy (20%)	Conf. Interval (95%)	Records

from disease								
All	0.959847	0.959847	0.04663718	0.7855781	0.08368305	0.9648087	0.1626824	12361
Train	0.960061	0.960061	0.04644928	0.7855781	0.0833953	0.9660194	0.1621313	8652
Test	0.9593695	0.9593695	0.04707549	0.7670119	0.08435045	0.9619844	0.1640254	3709
Valid	0.959847	0.959847	0.04663718	0.7855781	0.08368305	0.9648087	0.1626824	12361
Model 19								
Death from disease	R	Net-R	Avg. Abs.	Max. Abs.	RMS	Accuracy (20%)	Conf. Interval (95%)	Records
All	0.9706777	0.9706777	0.03681564	0.9505244	0.06690413	0.9847909	0.1300637	12361
Train	0.9727891	0.9727891	0.03624907	0.8932483	0.0642433	0.9867083	0.1248973	8652
Test	0.9659095	0.9659095	0.0381373	0.9505244	0.07273365	0.9803181	0.1414357	3709
Valid	0.9706777	0.9706777	0.03681564	0.9505244	0.06690413	0.9847909	0.1300637	12361
Model 20								
Death from disease	R	Net-R	Avg. Abs.	Max. Abs.	RMS	Accuracy (20%)	Conf. Interval (95%)	Records
All	0.9505929	0.9505929	0.04562325	0.9085212	0.08645324	0.9708206	0.1680654	14051
Train	0.9508882	0.9508882	0.04560892	0.8646238	0.08623064	0.9707168	0.1676402	9835
Test	0.9499053	0.9499053	0.04565668	0.9085212	0.08697029	0.9710626	0.1691118	4216

Probability survival curves were plotted over time for individual breast cancer patients with a composite curve illustrated in Figure 47 and individual curves illustrated in Figures 48 - 50.

Figure 47: Composite graph of predicted individual survival curves for different models compared to the actual survival.

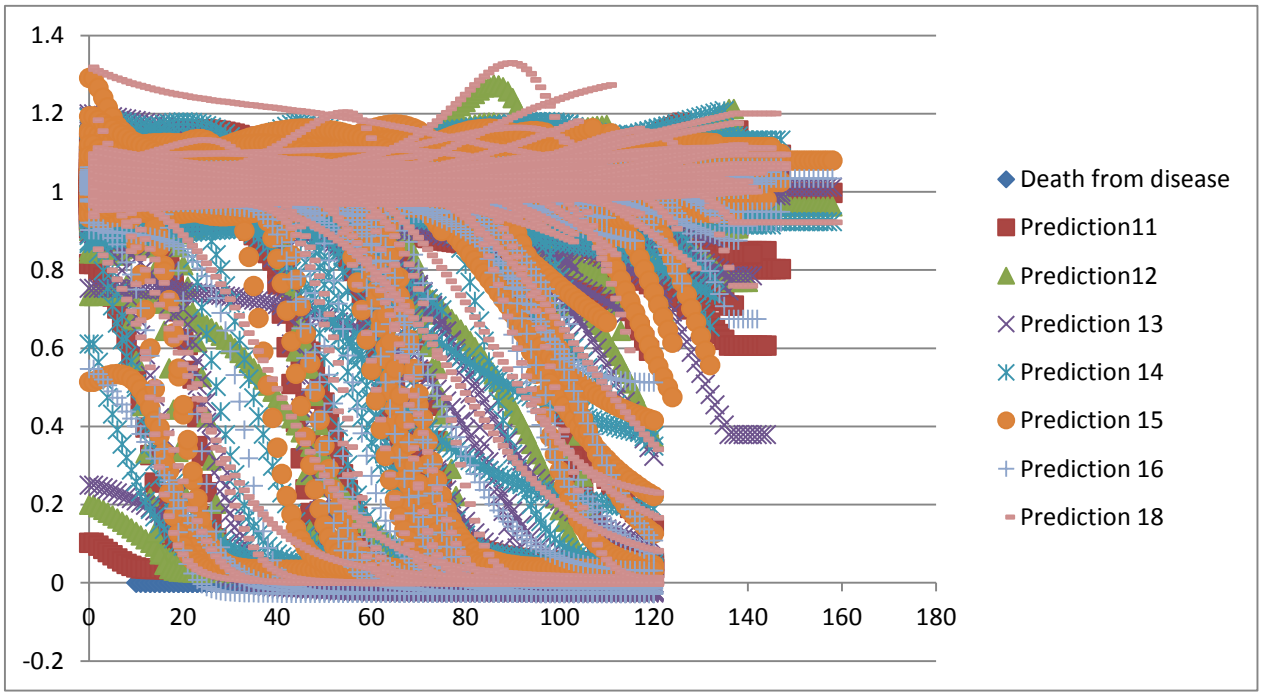


Figure 48: Probability survival curve for individual Patient comparing actual survival and predicted survival by ANN.

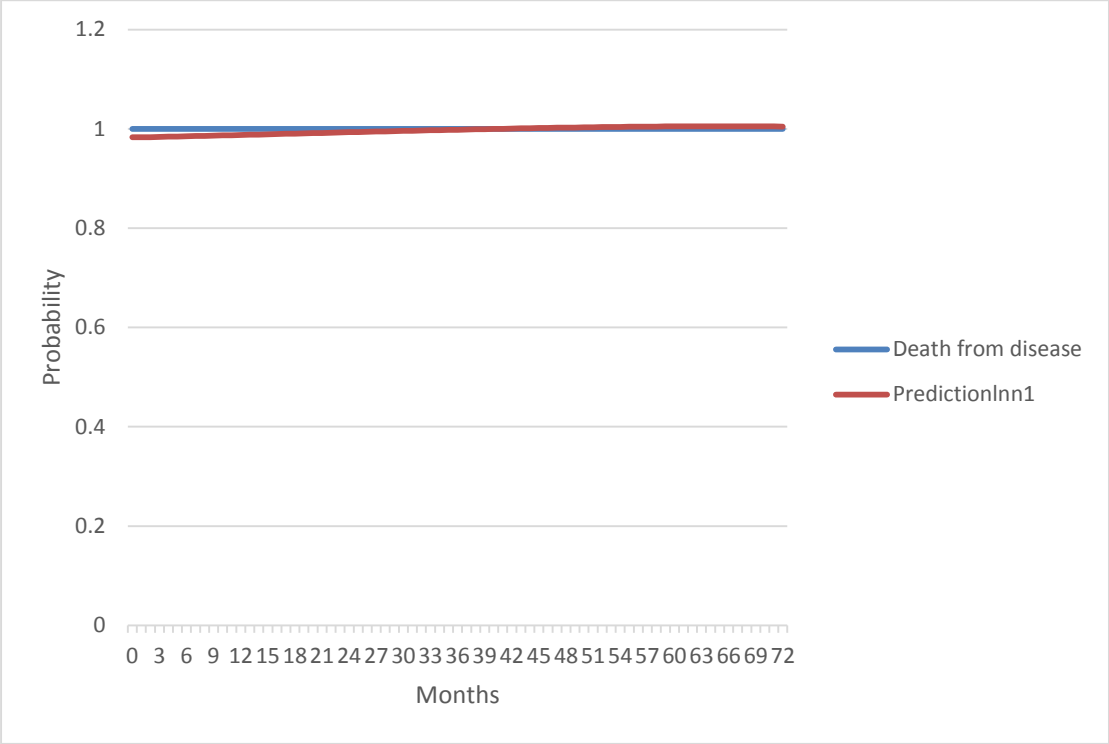


Figure 49: Probability survival curve for individual Patient comparing actual survival and predicted survival by ANN.

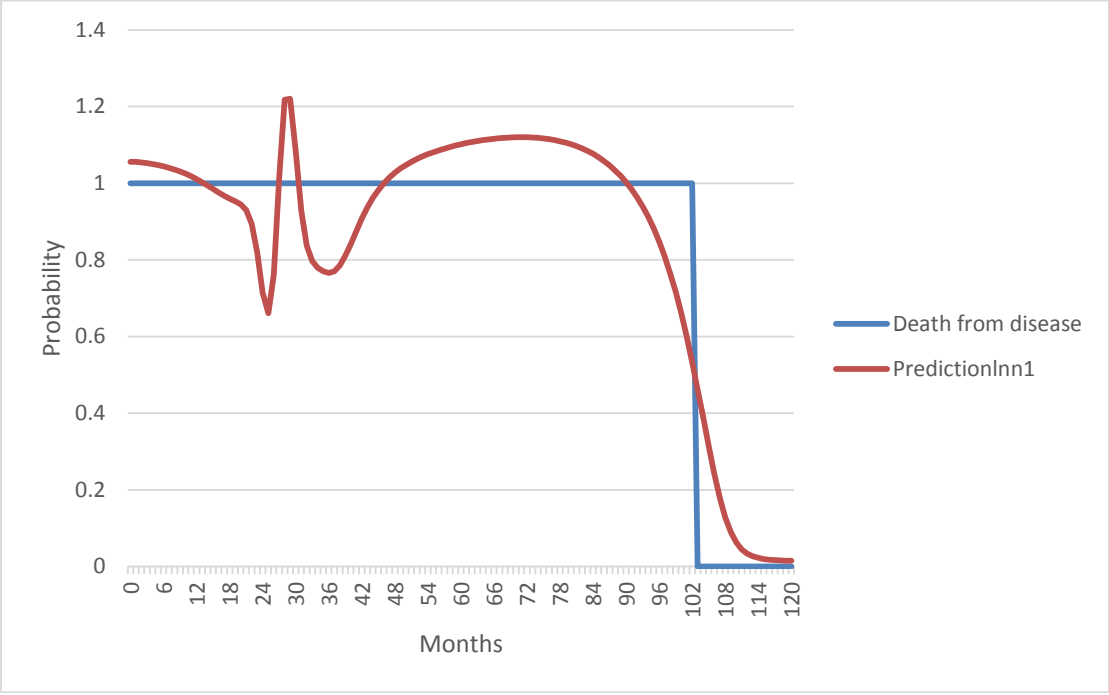


Figure 50: Probability survival curve for individual Patient comparing actual survival and predicted survival by ANN.

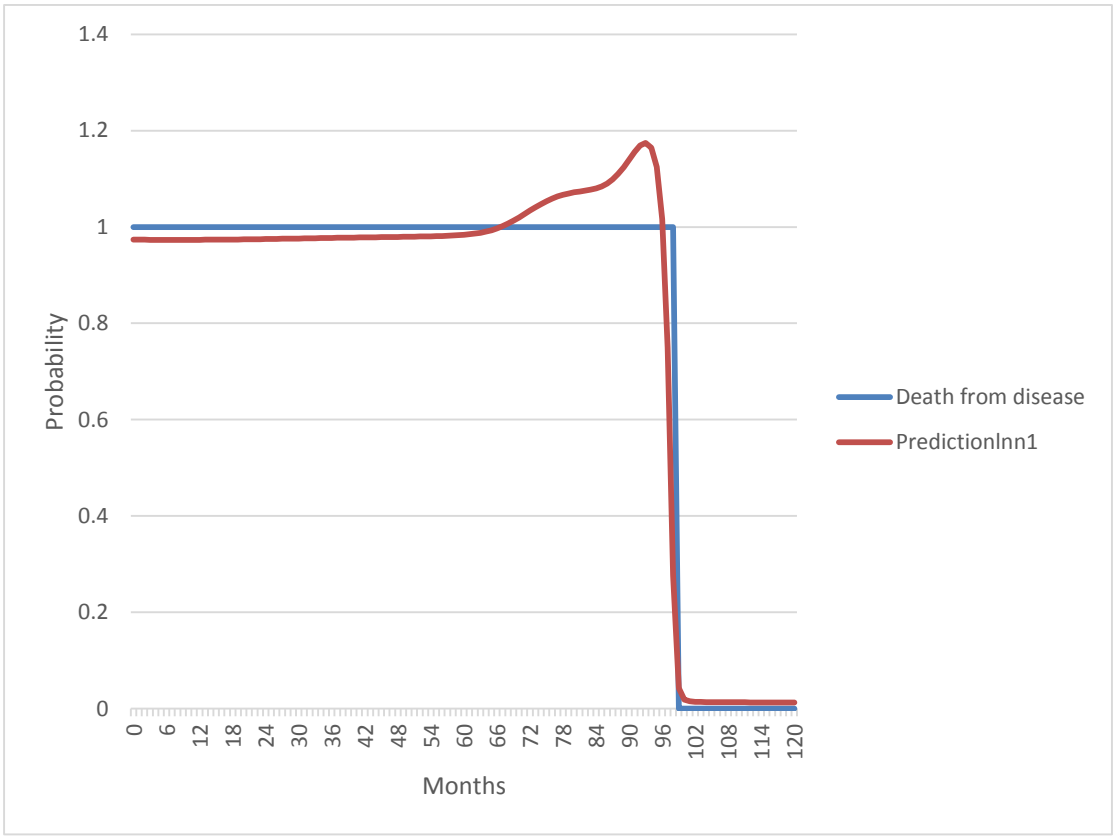
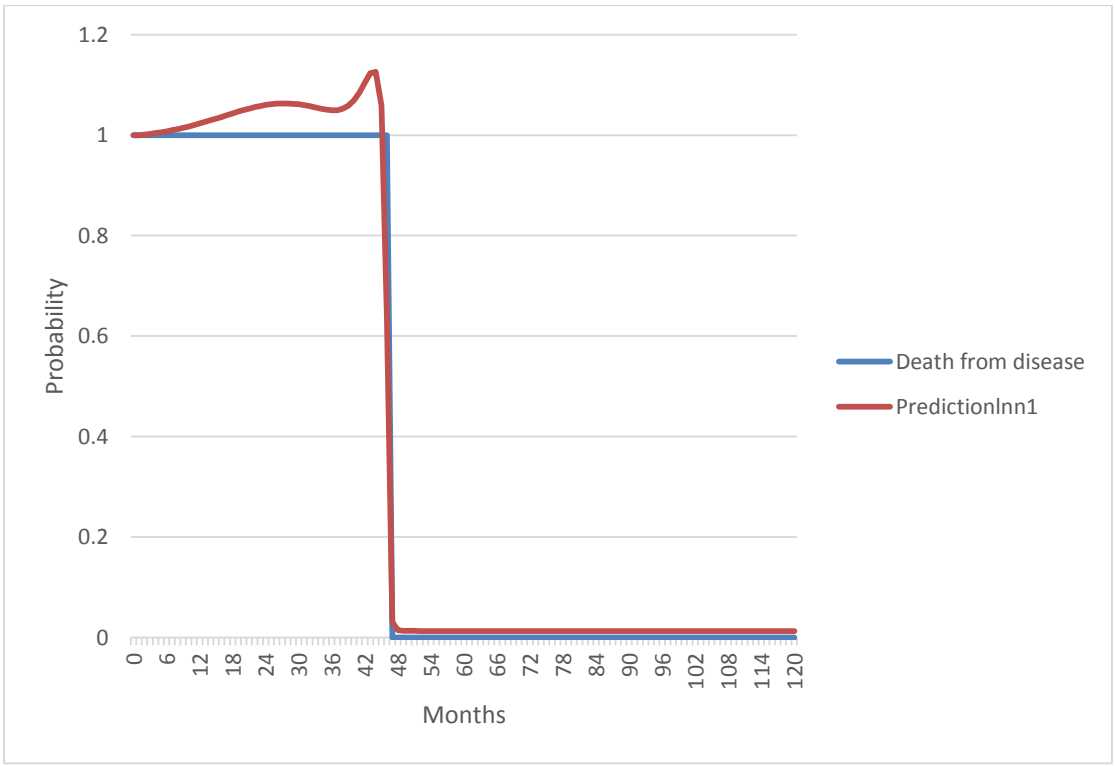


Figure 51: Probability survival curve for individual Patient comparing actual survival and predicted survival by ANN.

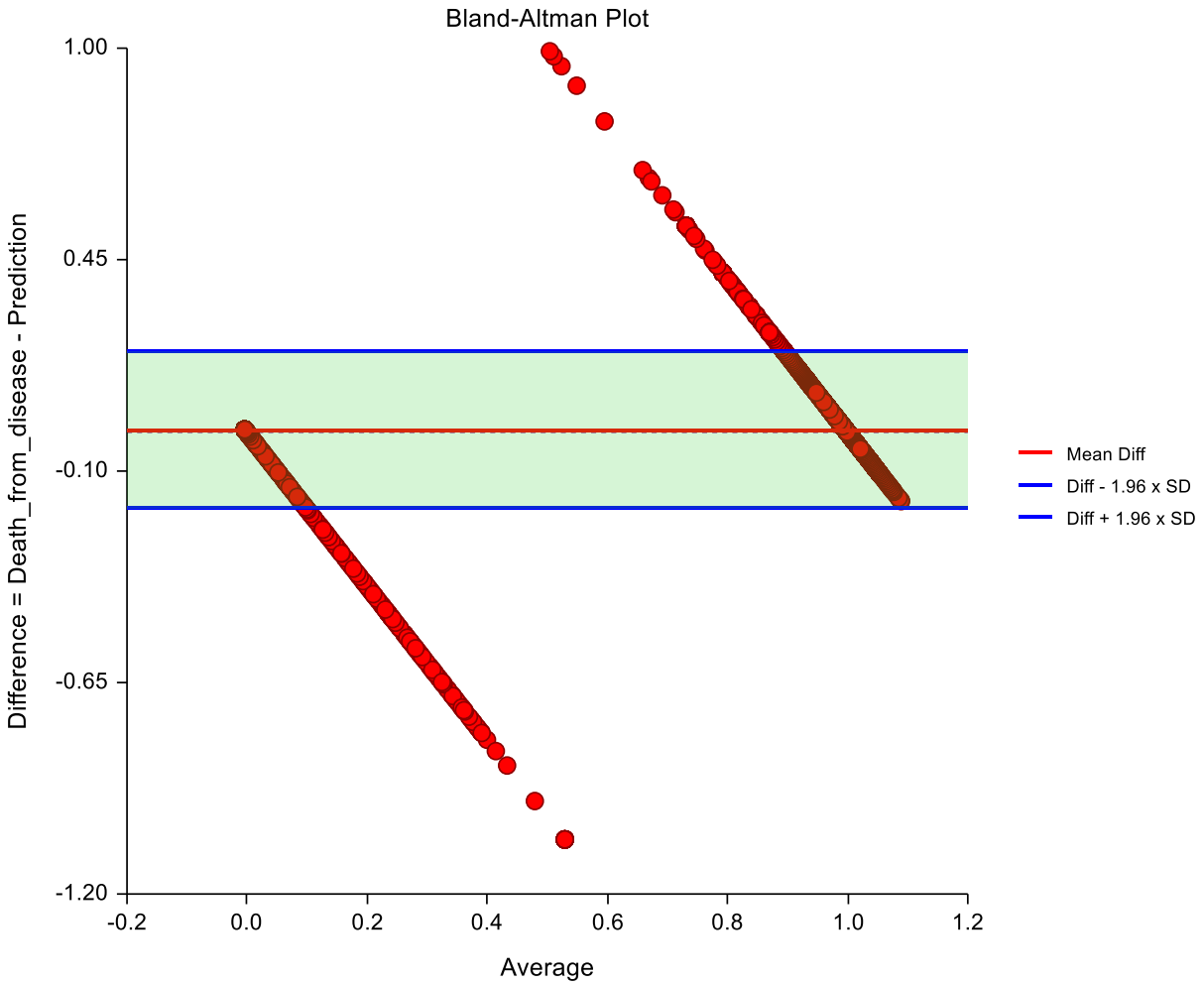


A Bland-Altman analysis was performed for comparison of the actual and predicted probabilities using NCSS 10 software for Model 1 (Figure 52).

Figure 52: Bland-Altman Plot showing correlation between BCSS (0=Deceased) and Predicted survival.

Bland-Altman Analysis Report

Bland-Altman Plot



Descriptive Statistics

Variable	Count	Mean	Standard Deviation	95.0% LCL of Mean	95.0% UCL of Mean
Death_from_disease	16127	0.9013456	0.2982067	0.896732	0.9059592
Prediction	16127	0.8945166	0.2926436	0.889989	0.8990441
Difference	16127	0.006829006	0.104128	0.005218022	0.00843999

Correlation Coefficient = 0.938055

Bland-Altman Analysis Report

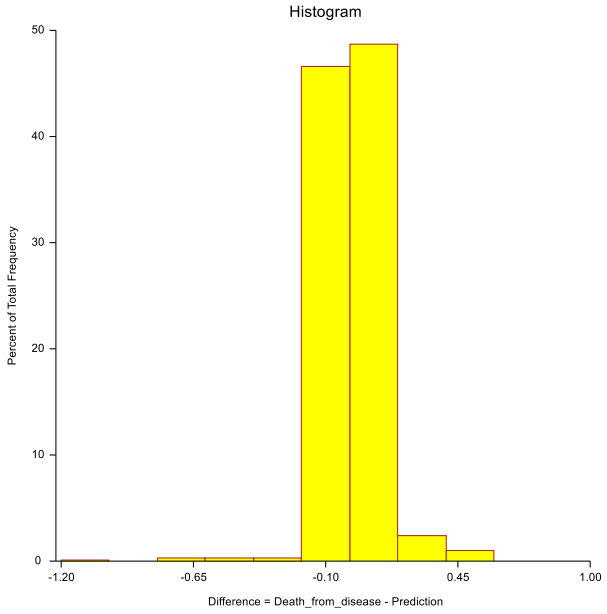
Bland-Altman Analysis: Bias and Limits of Agreement for Death_from_disease and Prediction
Limits of Agreement = Diff ± 1.96 x (Std Dev of Difference)

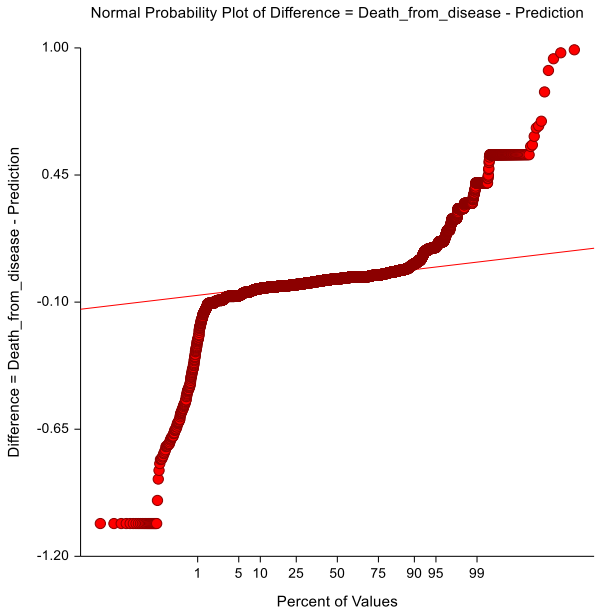
Parameter	Count	Value	Standard Deviation	95.0% LCL of Value	95.0% UCL of Value
Bias (Difference)	16127	0.006829006	0.104128	0.005218022	0.00843999
Lower Limit of Agreement	16127	-0.1972618	0.001401362	-0.2000084	-0.1945152
Upper Limit of Agreement	16127	0.2109198	0.001401362	0.2081732	0.2136664

Test of Normality of Differences Assumption

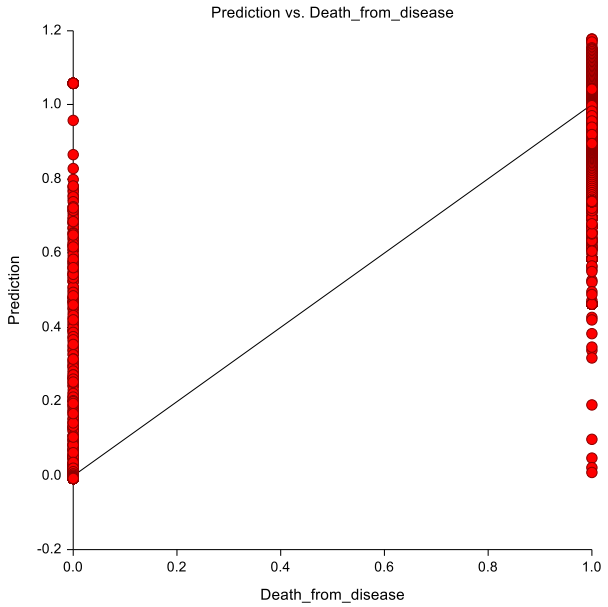
Assumption	Value	Prob Level	Decision (α = 0.050)
Shapiro-Wilk	0.595	0.0000	Reject normality

Evaluation of Assumptions Plots





Bland-Altman Analysis Report



13.6.3. Discussion

The ANN enabled development of a model to predict breast cancer specific survival in lymph-node negative individual patients over time using a combination of clinical features and biomarkers.